DOCUMENT RESUME

ED 203 299 CS 006 138

AUTHOR	Guthrie, John T., Ed.
TITLE	Comprehension and Teaching: Research Reviews.
INSTITUTION	International Reading Association, Newark, Del.
REPORT NO	ISBN-0-87207-943-0
PUB DATE	81
NOTE	333p.
AVAILABLE FROM	International Reading Association, 800 Barksdale Rd., P.O. Box 8139, Newark, DE 19711 (Order No. 943, $9.00 member, $13.50 non-member).
EDRS PRICE	MF01/PC14 Plus Postage.
DESCRIPTORS	Beginning Reading; Bilingual Education; *Cognitive Processes; *Reading Comprehension; Reading Difficulties; *Reading Instruction; *Reading Processes; Reading Programs; *Reading Research; Social Influences; Vocabulary Development
IDENTIFIERS	Schemata

ABSTRACT

Reflecting the interdisciplinary emphasis that reading comprehension has received during the past decade, the articles in this volume deal with both the processes involved in reading and the instructional practices used in teaching it. The six articles devoted to reading processes deal specifically with the following topics: schemata, comprehension of text structures, vocabulary knowledge, the social context of learning to read, and social-psychological perceptions and reading comprehension. The six articles concerning instructional practices report on instructional variables in reading comprehension, academic learning time and reading achievement, the role of reading in bilingual contexts, characteristics of exemplary reading programs, overcoming educational disadvantages, and recognizing reading comprehension programs. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED203299

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION
CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

- Points of view or opinions stated in this document do not necessarily represent official NIE position or policy

COMPREHENSION AND TEACHING:
RESEARCH REVIEWS

John T. Guthrie, Editor
International Reading Association

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

International
Reading Association

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS006138

ira International Reading Association
800 Barksdale Road • Newark, Delaware 19711

INTERNATIONAL READING ASSOCIATION

OFFICERS
1981 - 1982

President Kenneth S. Goodman, University of Arizona, Tucson

Vice President Jack Cassidy, Millersville State College, Pennsylvania

Vice President Elect Ira E. Aaron, University of Georgia, Athens

Executive Director Ralph C. Staiger, International Reading Association

DIRECTORS

Term Expiring Spring 1982
Nicholas P. Criscuolo, New Haven Public Schools, Connecticut
Bernice E. Cullinan, New York University, New York City
Patricia S. Koppman, San Diego Unified Schools, California

Term Expiring Spring 1983
Elizabeth Hunter-Grundin, Schools Council, London, England
Dianne L. Monson, University of Washington, Seattle
Alden J. Moe, Purdue University, West Lafayette, Indiana

Term Expiring Spring 1984
Phylliss Adams, University of Denver, Colorado
John Downing, University of Victoria, British Columbia, Canada
Sara Scroggins, St. Louis Public Schools, Missouri

Copyright 1981 by the
International Reading Association, Inc.

Library of Congress Cataloging in Publication Data
Main entry under title:

Comprehension and teaching.

Bibliography: p.
1. Reading comprehension—Addresses, essays, lectures. 2. Reading—Addresses, essays, lectures. 3. Comprehension—Addresses, essays, lectures. 4. Learning, Psychology of—Addresses, essays, lectures. I. Guthrie, John T.
LB1050.45.C65 428.4 80-21257
ISBN 0-87207-943-0

Contents

Foreword *v*

Introduction *vii*

PROCESSES IN READING

3 Schemata: The Building Blocks of Cognition
David E. Rumelhart

27 Comprehension of Text Structures
P. David Pearson and Kaybeth Camperell

56 On the Making of Inferences During Reading and Their Assessment *Thomas Trabasso*

77 Vocabulary Knowledge
Richard C. Anderson and Peter Freebody

118 Social Context of Learning to Read
Courtney B. Cazden

140 Social-Psychological Perceptions and Reading Comprehension
Herbert J. Walberg, Victoria Chou Hare, and Cynthia A. Pulliam

PRACTICES IN EDUCATION

163 Instructional Variables in Reading Comprehension
Joseph R. Jenkins and Darlene Pany

203 Academic Learning Time and Reading Achievement
David C. Berliner

227 The Role of Reading in Bilingual Contexts
Warwick B. Elley

255 Characteristics of Exemplary Reading Programs
S. Jay Samuels

274 Overcoming Educational Disadvantagedness
Harry Singer and Irving H. Balow

313 How to Recognize a Reading Comprehension Program: An Afterword *John T. Guthrie*

IRA PUBLICATIONS COMMITTEE 1981-1982 Robert B. Ruddell, University of California at Berkeley, *Chairing* • Phyllis J. Adams, University of Denver • Roach Van Allen, University of Arizona • Rosalinda B. Barrera, New Mexico State University • Janet R. Binkley, IRA • Faye R. Branca, IRA • MaryAnne Hall, Georgia State University • W. Dorsey Hammond, Oakland University • Martha L. King, Ohio State University • Lloyd W. Kline, IRA • Carol Kuykendall, Houston, Texas, Independent School District • Grover C. Mathewson, Florida International University • Joan Nelson, State University of New York at Binhamton • Lloyd O. Ollila, University of Victoria • P. David Pearson, University of Illinois at Champaign • María Elena Rodriguez, Asociacion Internacional de Lectura, Buenos Aires • S. Jay Samuels, University of Minnesota • Lyndon W. Searfoss, Arizona State University • Ralph C. Staiger, IRA • John C. Stansell, Texas A&M University • Marcella J. Taylor, Indiana Department of Public Instruction, Indianapolis • M. Jerry Weiss, Jersey City State College • Joanna Williams, Teachers College, Columbia University • Rosemary Winkeljohann, St. Ursula Villa Elementary School, Cincinnati.

The International Reading Association attempts, through its publications, to provide a forum for a wide spectrum of opinions on reading. This policy permits divergent viewpoints without assuming the endorsement of the Association.

Foreword

The chapters in this excellent volume reflect the strong interdisciplinary emphasis which reading comprehension has received during the past decade. The contribution of the cognitive scientist through schema theory, text structure, inferential reasoning, and word meaning is clearly in evidence. These efforts extend far beyond basic work of the past which focused heavily on the word and sentence isolated from natural language text. The impact of sociolinguistics is felt through the examination of instructional interaction in the language environment of the classroom. The importance of reading in bilingual education is not neglected in this collection and forecasts the great need for research in this area in the immediate years ahead.

A continued challenge is present in the reading field to bridge the chasm between understanding of the reading process and instructional practice. This area is not left untended. Emphasis is given to the importance of instructional time in the learning setting, to characteristics of exemplary reading programs, and to the impact of reading instruction spawned during a period of social unrest emphasizing the equality of educational opportunity.

This volume demonstrates the continued dedication of the International Reading Association to the improvement of literacy. John Guthrie and the contributors to this work are to be commended on their interdisciplinary and applied focus illuminating both reading process and practice.

Robert B. Ruddell
University of California at Berkeley

Introduction

During the 1970s, the scope of research on reading broadened. It now requires a larger definition to encompass studies that are designed to be about reading. A wider array of investigators seek to illuminate the many phenomena that characterize this human ability in its proficient stage, in its acquisition. There have always been investigators who use reading as a medium. Studies are often conducted on visual perception, language comprehension, persuasion, or consumerism, in which reading effectively serves as a tool. Although much can be learned about reading from this latter kind of research, the information is often inaccessible or buried through underinterpretation. It is the former collection that has broadened in definition and increased in number during the past decade.

To illustrate the research on reading that was prominent in the beginning of the 1970s, the annual summaries of reading, published in the *Reading Research Quarterly* in winter 1970, may be used. In that issue, 416 reports of reading research were compiled under William S. Gray's classic categories of the sociology, physiology, psychology, and teaching of reading. Within the psychology of reading, a preponderance of studies was on cognitive processes, although studies of language, personality, and sex differences made a nominal showing. The majority of published research papers on cognitive processes was focused on visual perception, auditory processes, and visual-auditory integration. The word, as opposed to the phrase, the paragraph, or the story, was used as the unit of analysis. Characteristics of words, such as their frequency of occurrence in written materials or the concreteness of nouns, were analyzed;

and paired associate learning was a favorite paradigm for studying their ease of acquisition. Research on these cognitive variables comprised 64 percent of the research on the psychology of reading. The only foreshadow of research to come in the '70s was the work of Lawrence Frase on questions and memory for text which was published mostly in the *Journal of Educational Psychology* and regarded at the time as intriguing but somehow beyond the pale of reading.

Also in 1970, the teaching of reading attracted the attention of educational researchers. Although a few fruitless statements about the status of reading instruction were being made (e.g., how many remedial reading teachers are located in certain districts of a state) and some forays into reading readiness were made, the bulk of the investigations pertained to methods of instruction. Of 25 studies on the teaching of reading, the comparison of allegedly different methods represents 80 percent of the group. This was the era of contrasting of initial teaching alphabet (i.t.a.) with conventional instruction; of visual perceptual training compared to no control; and basal instruction versus basal plus phonics. What typified these studies was the comparison of one or more series of tasks that were given to children to facilitate the acquisition of word recognition or reading comprehension ability. The experiments consisted of altering cognitive tasks or their sequences to examine the effect on reading achievement. Although a study by Labov on the relationship between reading achievement and school-related values of black adolescents was included in the annual summaries of 1970, this article was regarded as a contribution to sociology, rather than to our understanding of reading.

By 1980, the cognitive processes in reading under active investigation were expanded to include the comprehension of story structures, integration of sentences, drawing inferences, testing hypotheses, relating background knowledge to textual information, and reading as a process of information search. To accommodate a veritable explosion in these areas, several journals were founded, including *Cognitive Psychology*, *Discourse Processes*, and *Cognitive Science*. Some other periodicals have undergone a substantial reorientation. These veins are represented in the present volume in the chapters of David Rumelhart, Tom Trabasso, David Pearson, and Richard Anderson, who are not merely spokesmen but the agents of expansion in these topics.

It has been increasingly recognized that, although reading is a cognitive operation, learning to read is a social act. Since classrooms contain a large number of children who are often grouped into sections with different roles, responsibilities, expectations, activities, and relationships to the teacher, the social dimension of classrooms is coming to be a topic for close observation. Investigators who see classrooms as microsocieties and who use the observational tools of anthropologists have introduced some provocative concepts and perspectives to reading education. These are partially expressed in the present volume by Cazden and less directly by Walberg and his coauthors.

In studies of classroom practice for reading education, the boundaries have likewise been extended. Teaching reading has been redrawn to include not only a comparison between methods (which is here represented by Singer and Balow in a description of the Follow Through experiment), but also several independent strategies of investigation. One of these lines, summarized by Jenkins and Pany, consists of attempts to increase the proficiency with which children perform cognitive processes that are found to be related to reading ability in basic research efforts. A second type of study is focused on exemplary reading programs. Here summarized by Samuels, these investigations locate reading education programs known to be extraordinary for producing achievement and attempt to describe them as fully as possible. The expectation is that such descriptions may enable other educators to duplicate the programs elsewhere. Third, classroom practice in reading instruction has come to be viewed as a complex organization in which the analysis of time is essential. Berliner gives a summary of a common viewpoint about the importance of understanding and fostering the optimal use of time by teachers and pupils in the classroom to improve reading achievement.

A sobering set of problems in reading education stems from children whose languages or dialects are different from the "standard" used in schools. Although few coherent studies have addressed this problem within the United States, Elley reports that for children of the South Pacific, proficiency in oral language and reading can be best accomplished through simultaneous teaching toward an integrated set of goals.

What this volume intends to portray is the new topography in research on the processes and teaching of reading. To obtain the breadth of vision attempted here in a single volume, one necessarily reduces the clarity of focus. The findings of this book will provoke, not preclude, your verification; and the methodologies invite a cross-examination. Nevertheless, the authors of these chapters are to be commended for their openings into what we may come to see about reading.

JOHN T. GUTHRIE

Processes in Reading

Schemata: The Building Blocks of Cognition

David E. Rumelhart
University of California at San Diego

The notion of a schema and the related notions of frames, scripts, plans, etc., have been emphasized in Cognitive Science since the mid-1970s (cf. Bobrow & Norman, 1975; Chafe, 1976; Fillmore, 1975; Minsky, 1975; Moore & Newell, 1973; Rumelhart, 1975; Schank & Abelson, 1975; Winograd, 1975). It is my intent, in this paper, to introduce these concepts to readers unfamiliar with them and to show why so much attention has been paid to them. These various terms have been used by different authors to refer to any of a set of interrelated concepts. These terms are *not all synonymous*. Different authors have different things in mind when they use the different terms. Nevertheless, the various concepts are closely enough related that a discussion of any one of them will serve as an introduction to the others. I will thus focus my discussion on the one I know best, *schemata* (the singular is schema), as developed in Rumelhart and Ortony (1977).

The term schema comes into psychology most directly from Bartlett (1932). Bartlett himself attributes his use of the term to Head (1926). However, it would appear that Kant's (1787) use of the term already anticipated its major conceptual content. The OED gives the following definition of the term:

> In Kant: Any one of certain forms of rules of "productive imagination" through which the understanding is able to apply its "categories" to the manifold of sense-perception in the process of realizing knowledge or experience.

Adapted from a paper by the same title which appears in Spiro, Bruce, and Brewer (Eds.), *Theoretical Issues in Reading Comprehension*. Reprinted by permission of Lawrence Erlbaum Associates, Inc.

Some further discussion of Kant's view is given in Rumelhart and Ortony (1977). It is because of this historical precedence that I have chosen to retain the term schema.

For all of the authors mentioned above, schemata truly are *the building blocks of cognition.* They are the fundamental elements upon which all information processing depends. Schemata are employed in the process of interpreting sensory data (both linguistic and nonlinguistic), in retrieving information from memory, in organizing actions, in determining goals and subgoals, in allocating resources and generally in guiding the flow of processing in the system. Clearly, any device capable of all these wondrous things must be powerful indeed. Moreover, since our understanding of none of these tasks which schemata are supposed to carry out has reached maturity, it is little wonder that a definitive explication of schemata does not yet exist and that skeptics view theories based on them with some suspicion. In this paper, I hope to spell out, as clearly as possible, the nature of schemata and the kinds of problems they were devised to solve. In addition, I hope to present a convincing case that indeed the framework provided by schemata and allied concepts does, in fact, form the basis for a reasonable theory of human information processing.

My discussion through the next several sections of the paper will be abstract. Although I will not make direct application of these concepts to a theory of reading until near the end of the paper, many of the papers in this volume will illustrate the ways in which schemata can lead to insightful analyses of the reading process.

What Is a Schema?

A schema theory is basically a theory about knowledge. It is a theory about how knowledge is represented and about how that representation facilitates the *use* of the knowledge in particular ways. According to "schema theories" all knowledge is packaged into units. These units are the schemata. Embedded in these packets of knowledge is, in addition to the knowledge itself, information about how this knowledge is to be used.

A schema, then, is a data structure for representing the generic concepts stored in memory. There are schemata representing our knowledge about all concepts: those underlying objects, situations, events, sequences of events, actions and sequences of actions. A schema contains, as part of its specification, the network of interrelations that is believed to normally hold among the constituents of the concept in question. A schema theory embodies a *prototype* theory of meaning. That is, inasmuch as a schema underlying a concept stored in memory corresponds to the *meaning* of that concept, meanings are encoded in terms of the typical or normal situations or events which instantiate that concept.

Rather than attempting a formal description of schemata and their characteristics at this point, I will turn instead to some useful analogies in hopes of giving the reader a more concrete notion of the nature of schemata as I understand them. I will turn first to one of the more fruitful analogies, that of a play.

Schemata Are Like Plays

The internal structure of a schema corresponds, in many ways, to the script of a play. Just as a play has characters which can be played by different actors at different times without changing the essential nature of the play, so a schema has *variables* which can be associated with (bound to) different aspects of the environment on different instantiations of the schema. As an example, consider the schema for the concept *buy*. One can imagine a playwright having written a most mundane play in which the entire play consisted of one person purchasing some object from another person. At minimum, such a play must have two people, some merchandise, and some medium of exchange. Whatever else happens, at the outset of the play one character (call him/her the PURCHASER) must possess the medium of exchange (call it the MONEY). The second person, the SELLER must possess the object in question, the MERCHANDISE. Then, by some interaction (BARGAINING) a bargain is struck and the SELLER agrees to give the MERCHANDISE to the PURCHASER in exchange for a quantity of the MONEY. There would, of course, be many ways of playing this little play. The MERCHANDISE could vary from a trinket of little value to an object of incalculable worth. The SELLER and the PURCHASER could vary in status, occupation, sex, nationality, age, etc.; the MONEY could vary in amount, and whether it was actually money or clam shells; and the BARGAINING could vary in form. Still, through all of this variation, as long as the fundamental plot remained the same we could say that the BUY play was being performed.

Now, this little play is very much like the schema that I believe underlies our understanding of the concept *buy* or that for *sell*. There are variables corresponding to the characters in the play. We have the PURCHASER, the SELLER, the MONEY, the MERCHANDISE and the BARGAINING. When we understand a situation to be a case of BUYING, we come to associate persons, objects, and subevents with the variables of our schema. Upon having made these associations, we can determine to what degree the situation we are observing corresponds to this *prototype* case of BUYING.

Just as a playwright often specifies characteristics of the characters in his play (age, sex, disposition, etc.) so, too, as part of the specification of a schema, we have associated knowledge about the variables of the schema. We know, for example, that the PURCHASER and SELLER are normally people and that the MONEY is normally money. Moreover, we know that the value of the MONEY in question will co-vary with the value of the MERCHANDISE, etc. Such knowledge about the typical values of the variables and their interrelationships is called the *variable constraints*.

These constraints serve two important functions in a schema theory. In the first place, variable constraints help in the identification of the various aspects of the situation with the variables of the schema. If we know that we are observing a case of BUYING, we are not going to map the PURCHASER variable into the object in the world which should serve as the MONEY. We know this, in part, because we know that the PURCHASER is normally an animate being whereas the MONEY is normally money or some other inanimate object. In the second place, variable constraints can help by serving as *default values* (cf. Minsky, 1975) or initial "guesses" for variables whose values we have not yet observed. Thus, for example, if we take a certain transaction to be one of BUYING, but do not notice the MONEY, we can *infer* that there was MONEY and that, in fact, the MONEY probably *was money* amounting in value to about the value of the MERCHANDISE. In this way, the schema can help us make inferences about unobserved aspects of a situation.

It is perhaps useful to note here that variable constraints offer default values for unobserved variables *conditional on the values of the observed variables*. Moreover, the constraints are not *all-or-none* constraints which *require* that certain variables have a fixed range of values. Rather, they are merely specifications of the *normal* range of values for each variable and how this normal range varies with the specification of various combinations of other values on the other variables. Thus, as Rumelhart and Ortony (1977) suggest, it is perhaps most useful to think of variable constraints as forming a kind of *multivariate distribution* with correlations among the several variables.

There is also the notion of an *instantiation of a schema* which corresponds to an *enactment of a play*. A play is enacted whenever particular actors, speaking particular lines perform at a particular time and place. Similarly, a schema is *instantiated* whenever a particular configuration of values is bound to a particular configuration of variables at a particular moment in time. Interpreting a situation to be an instance of some concept, such as an instance of buying, involves, according to the present view, the instantiation of an appropriate schema, say the BUY schema, by associating the various variables of the schema with the various aspects of the situation. Such a schema along with its variable bindings is called an instantiated schema. Just as we could, say, take a movie of an enactment of a play and thereby save for posterity a trace of the enactment, likewise it is the traces of our instantiated schemata which serve as the basis of our recollections.

Before leaving the analogy between the script of a play and a schema, it is useful to note that neither is a complete specification of every detail—both allow room for irrelevant variation and creative interpretation. The script of a play, no matter how meticulous the playwright, allows for an infinity of variations, each of which can properly be considered an enactment of the play. Certain lines composed by the playwright are sometimes changed to suit the interpretation of the director. Nevertheless, within limits, it is the same play. So it is with schemata. A schema is not so rigidly applied that no variation is allowed. The schema only provides the skeleton around which the situation is interpreted. Variations orthogonal to the specifications of the schema have no bearing on the quality with which the schema is said to account for the situation. Moreover, even minor aspects of the situation which might be considered central to the schema can undergo some variation before we completely reject the interpretation provided by the schema.

Finally, despite all of the ways in which a schema is like a play, there are also numerous ways in which a schema is unlike a play. Perhaps most important of these is degree of abstraction. In our example of the BUY schema, we imagined a play that was more abstract than any one playwright would ever compose. Normally, the playwright would determine the *kind* of buying involved, as well as more detail about the characters and more constraints on the dialogue. The BUY schema, on the other hand, must be applicable to *any* case of buying and thus must, necessarily, be more abstract than any actual play would ever be. Moreover, whereas a play is normally about people and their actions, a schema may be about events and objects of any sort. Indeed, a schema may merely be about the nature of a wholly inactive object such as a chair. In this case, the schema specifies not action or event sequences, but rather spatial and functional relationships characteristic of chairs. Finally, although a play may contain acts, each with its own structure, a script for a play exists really only on one level. A script does not consist of a configuration of sub-scripts. A schema, on the other hand, should be viewed as consisting of a configuration of subschemata corresponding to the constituents of the concept being represented. These points will be made clearer in the following sections when I draw analogies between schemata and other familiar concepts.

Schemata Are Like Theories

Perhaps the central function of schemata is in the construction of an interpretation of an event, object, or situation—that is, in the process of comprehension. In all of this, it is useful to think of a schema as a kind of informal, private, unarticulated theory about the nature of the events, objects, or situations which we face. The total set of schemata we have available for interpreting our world in a sense constitutes our private theory of the nature of reality. The total set of schemata instantiated at a particular moment in time constitutes our internal model of the situation we face at that moment in time. Or, in the case of reading a text, a model of the situation is depicted by the text.

Thus, just as the activity surrounding a theory is often focused on the evaluation of the theory and the comparison of the theory with observations we have made, so it is that the primary activity associated with a schema is the determination whether it gives an adequate account for some aspect of our current situation. Just as the determination that a particular theory accounts for some observed results involves the determinations of the *parameters of the theory*, so the determination that a particular configuration of schemata accounts for the data presently available at our senses requires the determination of the values of the *variables of the schemata.* If a promising schema fails to account for some aspect of a situation, one has the options of accepting the schema as adequate in spite of its flawed account or of rejecting the schema as inadequate and looking for another possibility. Therefore, the fundamental processes of comprehension are taken to be analogous to hypothesis testing, evaluation of goodness of fit and parameter estimation. Thus, a reader of a text is presumably constantly evaluating hypotheses about the most plausible interpretation of the text. Readers are said to have understood the text when they are able to find a configuration of hypotheses (schemata) which offer a coherent account for the various aspects of the text. To the degree to which a particular reader fails to find such a configuration, the text will appear disjointed and uncomprehensible.

Schemata are like theories in another important respect. Theories, once they are moderately successful, become a source of predictions about unobserved events. Not all experiments are carried out. Not all possible observations are made. Instead, we use our theories to make inferences with some confidence about these unobserved events. So it is with schemata. We need not observe all aspects of a situation before we are willing to assume that some particular configuration of schemata offers a satisfactory account for that situation. Once we have accepted a configuration of schemata, the schemata themselves provide a richness which goes far beyond our observations. Upon deciding that we have seen an automobile, we assume that it has an engine, headlights, and all of the standard characteristics of an automobile. We do this without the slightest hesitation. We have complete confidence in our little theory. This allows our interpretations to far outstrip our sensory observations. In fact, once we have determined that a particular schema accounts for some event we may not be able to determine which aspects of our beliefs are based on direct sensory information and which are merely consequences of our interpretation.

Schemata Are Like Procedures

There are at least two inadequacies of the analogies presented above. In the first place, plays and theories are passive. Schemata are active processes. In the second place, the relationship between a theory and its constituent sub-theories or between a play and its constituent sub-plays is not always evident. Schemata, on the other hand, have a very well defined constituent structure.

In both of these ways, schemata resemble procedures or computer programs. Schemata are active computational devices capable of evaluating the quality of their own fit to the available data. That is, a schema should be viewed as a procedure whose function it is to determine whether, and to what degree, it

18

accounts for the pattern of observations. This includes, among other things, associating its variables to the appropriate aspects of its environment—i.e., binding its own variables. Thus, to the degree that schemata underlying concepts are identified with *meaning of those concepts*, a schema theory is *both* a *prototype theory* and a *procedural theory* of meaning. Obviously, the degree to which a schema theory of human information processing can work depends on the degree to which procedures can actually be constructed to carry out the tasks I have just assigned to them. I believe they can and will address this issue in the following sections.

The second characteristic which schemata share with procedures is a structural one. Procedures normally consist of a network (or a tree) of sub-procedures. A particular procedure normally carries out its task by invoking a pattern of sub-procedures each of which in turn operates by invoking its sub-procedures. Each procedure or sub-procedure can return values which can serve as conditions determining which if any further sub-procedures are to be invoked. So it is with schemata. A schema is a network (or possibly a tree) of sub-schemata each of which carries out its assigned task of evaluating its goodness of fit whenever activated. These sub-schemata represent the conceptual constituents of the concept being represented.

Thus, for example, suppose we had a schema for a FACE. This would consist of a certain configuration of sub-schemata each representing a different constituent of a face. For example, there would presumably be a sub-schema representing the MOUTH, one for the NOSE and one for each EAR and each EYE. These sub-schemata in turn would consist of a configuration of constituents. The EYE schema, for example, would consist of a configuration of sub-schemata including, perhaps, an IRIS, EYELASHES, an EYEBROW, etc.

Just as a procedure uses results produced by its sub-procedures to carry out its task, so too a schema uses results produced by its sub-schemata to carry out its tasks. As I indicated above, the primary activity of a schema is the evaluation of its goodness of fit. An important mechanism of this evaluation involves the evaluation of the goodness of fit of each of its constituent parts. Thus, if a good EYE is found and a good MOUTH is found, the FACE schema can use this information along with its own evaluation of whether the entire *configuration* is right for a face to generate an overall evaluation of its goodness of fit.

To summarize then, just as a procedure consists of sub-procedures and those sub-procedures, in turn, consist of more sub-procedures, etc., so a schema consists of sub-schemata each of which, in turn, is specified as a configuration of its sub-schemata, etc. One may be struck by the fact that this process must stop somewhere. If each and every schema were merely a configuration of sub-schemata the process would never end. The solution to this dilemma for schemata is identical to the solution for procedures. When a computer program is written, this embedding process does not continue indefinitely. Eventually, some sub-procedure consists entirely of a configuration of *elementary instructions* for the machine in question. Likewise, with schemata, there must be a set of schemata which are elementary, in the sense that they do not consist of a further breakdown in terms of sub-schemata. Such elementary schemata correspond to what Norman and Rumelhart (1975) call *primitives.*

Schemata Are Like Parsers

A parser is a device which, given a sequence of symbols, determines whether that sequence forms a legal sentence (according to the rules of some grammar) and, if it does, determines the *constituent structure* of the sentence. That is, it determines which symbols in the sequence correspond to which constituents of the sentence. The process of finding and verifying appropriate schemata is thus a kind of parsing process which works with conceptual elements—finding constituents and sub-constituents among the data currently impinging on the system in much the same way that a sentence parser must find the proper parse for the input string of words.

One particularly useful aspect of this analogy is the substantial body of work carried out in computational linguistics on various parsing procedures. I believe that the processing strategies developed for some of the most sophisticated of these will carry over nicely in their application to schemata generally. As I will discuss below, I have in mind here especially the work of Kaplan (1973) and his development of the general syntactic processer (GSP).

Summary of the Major Features of Schemata

Rumelhart and Ortony (1977) listed four major characteristics of schemata. These were:

1. Schemata have variables.
2. Schemata can be embedded, one within another.
3. Schemata represent knowledge at all levels of abstraction.
4. Schemata represent *knowledge* rather than definitions.

The analogies presented above illustrate all of these features. Whereas schemata have variables, plays have roles, theories have parameters, and procedures have arguments. The embedding characteristic of schemata is best illustrated by the analogy between schemata and procedures. Schemata consist of subschemata as procedures consist of sub-procedures. Just as theories can be about the grand and the small, so schemata can represent knowledge at all levels—from ideologies and cultural truths to knowledge about what constitutes an appropriate sentence in our language to knowledge about the meaning of a particular word to knowledge about what patterns of excitations are associated with what letters of the alphabet. We have schemata to represent all levels of our experience, at all levels of abstraction. Finally, our schemata *are* our knowledge. All of our generic knowledge is embedded in schemata.

In addition to these four features, the analogies presented here indicate at least two more general features of schemata:

5. Schemata are active processes.
6. Schemata are recognition devices whose processing is aimed at the evaluation of their goodness of fit to the data being processed.

The Control Structure of Schemata

Perhaps the central questions in the development of a schema based model of perception and comprehension are: first, how is an adequate configuration of schemata discovered and, second, how is the goodness of fit evaluated? These are largely problems of *control structures*. There are many schemata. Not all of them can be evaluated at once. Somehow, there must be a schema for activating just those schemata which are most promising. There are two basic sources of activation for schemata. These are usually referred to as *top-down* and *bottom-up* activation. These two directions correspond to what Bobrow and Norman (1975) have called *conceptually-driven* and *data-driven* processing. I turn now to a discussion of these two modes of activation.

Conceptually-Driven and Data-Driven Processing

A schema may activate a sub-schema in the way a procedure invokes its sub-procedures. This is called *conceptually-driven* processing. In a sense, conceptually-driven processing is *expectation-driven* processing. That is, when a schema is activated and it, in turn, activates its sub-schemata, the activation of these sub-schemata are derived from a sort of expectation that they will be able to account for some portion of the input data. For example, suppose that, through some mechanism, the FACE schema is considered a promising account for the input and thereby activated and set about evaluating its goodness of fit. The *promise* of the FACE schema is, in a sense, transferred to its MOUTH, NOSE, EYE, EAR, etc. sub-schemata.

A second mechanism for schema activation is bottom-up or *data-driven* activation. A schema is said to be activated from the bottom-up whenever a sub-schema which has been somehow activated causes the various schemata of which it is a part to be activated. If the activation of the FACE schema led to the activation of the PERSON schema, we would say that the activation of the PERSON schema was data-driven. Thus, where *conceptually-driven* activation goes from *whole to part*, *data-driven* activation goes from *part to whole*. In schema directed processing, activation goes in *both* directions.

Schema directed processing is assumed to proceed in roughly the following way: Some event occurs at the sensory system. The occurence of this event "automatically" activates certain "low-level" schemata (such schemata might be called *feature detectors*). *These* low level schemata would, in turn, activate (in a data-driven fashion) certain of the "higher level" schemata (the most probable ones) of which they are constituents. These "higher level" schemata would then initiate conceptually-driven processing by activating the sub-schemata not already activated in an attempt to evaluate its goodness of fit.

At some point, when one of these higher level schemata began to get further positive results about its goodness of fit (i.e., it found evidence for other of its constituents), it would activate still higher level schemata which would look for still larger constituents.

This higher, more abstract schema would then activate, from the top-down, still other of its constituent schemata; and this activation would flow through its sub-schemata back down to lower-level schemata which would eventually make contact with either other schemata which have been activated from the bottom-up or they will initiate a search for the "predicted" sensory inputs.

Whenever a schema initiates a search for sensory data which are not present, that counts as evidence against that schema and also as evidence against all of those schemata which require the presence of that schema as a constituent sub-schema. When sufficient evidence is accumulated against a schema, processing of that schema is suspended and processing resources are allocated to other currently more promising schemata. Whenever enough evidence is gained in favor of a schema, that schema is taken as an adequate account for the relevant aspect of the input and the interpretation offered by that schema is taken as the "correct" interpretation of the relevant event. Later processing, on other, higher-level schemata may eventually disconfirm a temporarily accepted schema and we will have the phenomenon of the "double-take."

My discussion of the processing system to this point has been rather abstract. In the following section I will examine, in some detail, an example of this mixed initiative processing system.

An Example

Consider the following brief passage:

> Business had been slow since the oil crisis. Nobody seemed to want anything really elegant anymore. Suddenly the door opened and a well dressed man entered the showroom floor. John put on his friendliest and most sincere expression and walked toward the man.

Although merely a fragment, most people generate a rather clear interpretation of this story. Apparently, John is a car salesman fallen on hard times. He probably sells rather large, elegant cars (most likely, Cadillacs). Suddenly a good prospect enters the showroom where John works. John wants to make a sale. To do that he must make a good impression on the man. Therefore he tries to appear friendly and sincere. He also wants to talk to the man to deliver his sales pitch. Thus, he makes his way over to the man. Presumably, had the story continued John would have made the sales pitch and, if all went well, sold the man a car.

How do people arrive at such an interpretation? Clearly, people do not arrive at it all at once. As the sentences are read, schemata are activated, evaluated, and refined or discarded. When people are asked to describe their various hypotheses as they read through the story, a remarkably consistent pattern of hypotheses generation and evaluation emerges. The first sentence is usually interpreted to mean that business is slow *because* of the oil crisis. Thus, people are led to see the story as about a business which is somehow dependent on oil and is suffering. Frequent hypotheses involve either the selling of cars, or of gasoline. A few interpret the sentence as being about the economy in general. The second sentence, about people not wanting elegant things anymore, leads people with the gas station hypothesis into a quandary. Elegance just doesn't fit with gas stations. The gas station hypothesis is weakened, but not always rejected. On the other hand, people with hypotheses about the general economy or about cars have no trouble incorporating this sentence into their emerging interpretation. In the former case, they conclude it means that people don't buy luxury items; and, in the latter, they assume it means that people don't buy large elegant cars—Cadillacs—much anymore. The third sentence clinches the car interpretation for nearly all readers. They are already looking for a business interpretation—that most probably means a SELLING interpretation—and when a *well-dressed man* enters the door he is immediately labeled as someone with MONEY—a prospective BUYER. The phrase *showroom floor* clearly invalidates the gas station interpretation and strongly implicates automobiles which are often sold from a showroom. Moreover, the occurrence of a specific event doesn't fit at all well with the view that the passage is a general discussion of the state of the economy. Finally, with the introduction of John, we have an ideal candidate for the SELLER. John's actions are clearly those stereotypic of a salesman. John wants to make a sales and his "putting on" is clearly an attempt on his part to "make a good impression." His movement toward the man fits nicely into this interpretation. If he is a salesman, he must make contact with the man and deliver the stereotypic "pitch."

Qualitatively, this little account (which was derived from an analysis of a number of readers describing their current interpretation of the story after each sentence) fits well with the general approach I have been outlining. The process of comprehension is very much like the process of constructing a theory, testing it against the data currently available, and, as more data become available, specifying the theory further—i.e., refining the default values (as perhaps was the case when those holding the "car hypothesis" from the beginning encountered the sentence about nobody wanting anything elegant anymore). If the account becomes sufficiently strained, it is given up and a new one constructed, or, alternatively, if a new theory presents itself which obviously gives a more cogent account, the old one can be dropped and the new one accepted.

But where do these theories come from? The theories are, of course, schemata. Presumably, through experience we have built up a vast repertoire of such schemata. We have schemata for salesmen, the kinds of motives they have and the kinds of techniques they employ. We have schemata for automobiles, including how and where they are sold. We have built up schemata for the "oil crisis," what kinds of effects it has on what kinds of businesses. We have schemata about business people, the kinds of motives they have and the kinds of responses they make to these motives. The knowledge embedded in these schemata forms the framework for our theories. It is some configuration of these schemata which ultimately forms the basis for our understanding.

But how does a relevant schema suggest itself? It is here that the control structures discussed above play an essential role. Presumably, it is the bottom-up observation that a certain concept has been referenced that leads to the suggestion of the initial hypotheses. The notion that business was slow, suggests schemata about business and the economy. Since the slowness was dated from the occurrence of the oil crisis, it is a natural inference that the oil crisis was the *cause* of the slowness. Thus, a BUSINESS schema is activated. The particular TYPE of business is presumably a variable which must be filled. The information about the oil crisis suggests that it may be an oil related business. Thus, readers are led to restrict the TYPE variable of the BUSINESS schema to oil related businesses.

At this point, after the bottom-up activation of the high level BUSINESS schema has occurred, this schema would generate a top-down activation of the various possible oil related businesses. Prime candidates for these are, of course, automobile related businesses. Of these, selling gasoline and automobiles are the two most salient possibilities.

When the second sentence is encountered, an attempt is made to fit it into the schemata currently considered most promising. As I discussed above, this information could serve to further restrict the TYPE variable in the automobile BUSINESS schema but doesn't fit well with the gasoline business schema.

The BUSINESS schema presumably has, as part of its specification, a reference to the BUY or SELL schema discussed earlier. Once activated, these schemata search for potential variable bindings. In the case of the automobile business, the MERCHAN-

DISE variable is bound to an automobile. The second sentence suggests an elegant automobile. When the third sentence is encountered, the reader has not yet found a candidate for BUYER or SELLER. The sentence about a well-dressed man immediately suggests a potential BUYER. The phrase "showroom floor" offers additional bottom-up support for the automobile hypothesis. In fact, it is a strong enough clue itself that it can suggest automobile sales to a reader who currently considers an alternative schema more likely. We thus have a BUYER and some MERCHANDISE. The well-dressed quality of the BUYER is consistent with our view that the MERCHANDISE is elegant and therefore expensive—being well-dressed suggests MONEY. We need only a SELLER—i.e., an automobile salesman. Readers probably already bring a relatively complete characterization of the "default value" for car salesman. We need but little additional information to generate a rather detailed description of his goals and motives.

In spite of the length of this example, it should be noted that I have provided only a sketch of the elaborate processing which must occur in the comprehension of even so simple and direct a story as this. The problem is indeed a complex one and no one yet has been able to construct a model capable of actually carrying out the tasks involved. The conviction that the concept of the *schema* is the most promising route to the solution to these problems has led to its current popularity.

The Major Functions of Schemata

My intent to this point has been primarily definitional. I have tried to show what schemata are and generally how they are supposed to work. In this section I will give a few examples, mostly taken from the psychological literature, of phenomena for which schemata appear to offer promising accounts. I first turn to a discussion of *perception,* especially as it relates to reading.

Schemata and Perceiving

There are numerous examples in the psychological literature which suggest a schema-like theory to account for them. I will mention just a few examples here. Perception, like languag comprehension, is an interactive process. Information comes in from our sense organs, which suggest but do not determine appropriate schemata for the interpretation of the sense data. It is often only in the context of the whole that the individual parts of an object can be identified. Similarly, the whole itself cannot be identified apart from its parts. The interpretation of parts and wholes must proceed jointly. Our final interpretation is determined both by the local clues and by consistency among the various levels of analysis. Consider, as an example, Figure 1 taken from Palmer (1975). The object on the left is clearly recognizable as a face, but its parts (series B) are not recognizable out of context. Thus, it cannot be that we first perceive the parts and then construct an interpretation of the whole. Rather, the various shapes of the lines *suggest*, but do not determine, possible interpretations (the wiggly line suggests a possible nose, the acute angle suggests a possible eye, etc.). Lower level NOSE and EYE schemata may be activated, which in turn may activate higher level schemata such as the FACE schema. The FACE schema then activates schemata for all of the parts of the FACE not receiving bottom-up activation. (For example, the lips may not be close enough to LIPS to activate this schema at all out of context. In this case, the LIPS schema would be activated by the FACE schema and find sufficient evidence to serve—in context—to count as LIPS.)

As can be noted from series C of the Figure, it is not that parts of a face cannot *ever* be recognized without the face as a context. But, in order to be recognized out of context, they too must have an internal structure. If enough data is available about its internal structure, a schema like the NOSE schema can serve the function of an organizing whole perfectly well.

There is ample evidence of similar processes in the reading process. It is well known, for example, the strings of characters which form words are more easily apprehended than strings which do not form words. The reason for this presumably stems from the fact that we have schemata corresponding to words and none for random letter strings. Just as evidence for a NOSE indirectly constitutes evidence for LIPS through the FACE schema, so too evidence for one letter can constitute evidence for other letters through the schema for the word in question. Thus, for ex-

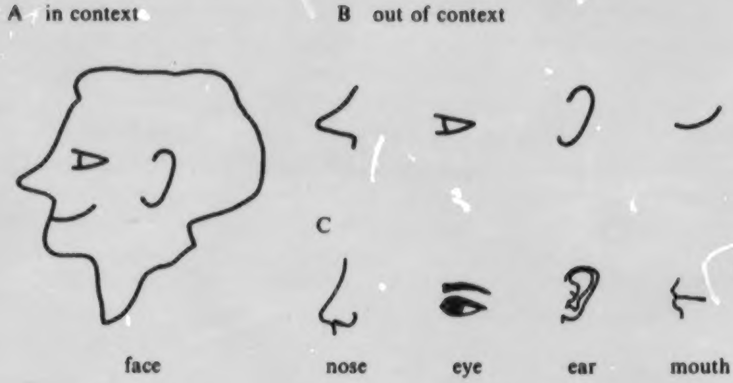


Figure 1
An illustration of part-whole context. Facial features recognizable in the context of a profile (A) are not recognizable out of context (B). When the internal part structure of the facial features is differentiated (C), however, the features become recognizable out of context.

ample, evidence favoring a T in the first position and an E in the third position of a three letter word indirectly constitutes evidence for a H in the second position through activation of the THE schema. The use of such information is presumably the mechanism whereby words are easier to see than random letter strings. Moreover, one of the characteristics which separate skilled readers from those with less skill is presumably the availability of more, more completely developed word schemata.

It is interesting that schemata not only *contribute* towards the development of an accurate perception but, by the same token they can sometimes cause a distortion. An experiment by Bruner and Potter (1964) illustrates the debilitating effect of premature commitment to a particular schema. In the study, subjects were presented with unfocused slides of familiar objects. The slides were slowly brought into focus. At each step along the way, as the slides were brought into focus, subjects were to report their best guess as to what the content of the slide was. Under these conditions, subjects continued to mis-identify the object long after naive subjects (those started with less severe amounts of defocusing) were able to readily identify the object in question.

This result is presumably due to the fact that subjects became committed to their early interpretations of the slide and then needed more information to disconfirm their original hypothesis than is normally required.

Schemata and Understanding Discourse

As discussed above, the process of understanding discourse is the process of finding a configuration of schemata which offers an adequate account of the passage in question. The analysis of the "oil crisis story" given above illustrates generally how such a process is supposed to operate. Clues from the story suggest possible interpretations (instantiations of schemata) which are then evaluated against the successive sentences of the story until finally a consistent interpretation is discovered. Sometimes, a reader fails to correctly understand a passage. There are at least three reasons implicit in schema-theory as to why this might occur. 1) The reader may not have the appropriate schemata. In this case he/she simply cannot understand the concept being communicated. 2) The reader may have the appropriate schemata, but the clues provided by the author may be insufficient to suggest them. Here again the reader will not understand the text but, with appropriate additional clues, may come to understand it. 3) The reader may find a consistent interpretation of the text, but may not find the one intended by the author. In this case, the reader will "understand" the text, but will misunderstand the author.

There are numerous examples of these three phenomena in the literature. Perhaps the most interesting studies along these lines were carried out by Bransford and Johnson (1973). They studied the comprehension of texts in which subjects lacked the appropriate schemata, ones in which the schemata were potentially available, but there were not sufficient clues to suggest the correct ones as well as ones on which subjects were led to choose a "wrong" interpretation. Consider, as an example, the following paragraph used in one of the studies.

> The procedure is actually quite simple. First you arrange things into different groups. Of course, one pile may be sufficient depending on how much there is to do. If you have to go somewhere else due to lack of facilities that is the next step, otherwise you are pretty well set. It is important not to overdo things. That is, it is better to do too few things at once than too many. In the short run this may not seem important but complications can easily arise. A mistake can be expensive as well. At first the whole procedure will seem complicated. Soon, however, it will become just another facet of life. It is difficult to foresee any end to the necessity for this task in the immediate future, but then one can never tell. After the procedure is completed one arranges the materials into different groups again. Then they can be put into their appropriate places. Eventually they will be used once more, and the whole cycle will then have to be repeated. However, that is part of life. [p. 400]

Most readers find this passage extremely difficult to understand. However, once they are told that it is about washing clothes, they are able to bring their clothes washing schema to the fore and make sense out of the story. The difficulty with this passage is thus not that readers don't have the appropriate schemata; rather, it stems from the fact that the clues in the story never seem to *suggest* the appropriate schemata in the first place. The "bottom-up" information is inadequate to initiate the comprehension process appropriately. Once the appropriate schemata are suggested, most people have no trouble understanding the text.

Although most readers simply find the passage incomprehensible, some find alternative schemata to account for it and thus render it comprehensible. Perhaps the most interesting interpretation I have collected was from a Washington bureaucrat who had no difficulty with the passage. He was able to interpret the passage as a clear description of his job. He was, in fact, surprised to find that it was supposed to be about "washing clothes" and not about "pushing papers." Here, then, we have an example of the third kind of comprehension failure, "understanding the story" but "misunderstanding the author."

Obviously, a detailed account of the comprehension process requires a detailed description of the schemata readers have available, as well as an account of the conditions under which certain of these schemata are activated. A number of researchers have been developing such specific models of specific schemata (cf. Rumelhart, 1975, 1977; Schank & Abelson, 1975). Other investigators have described how schemata are used to understand narratives (Mandler & Johnson, 1977), expository texts (Meyer, 1975), and newspaper articles (Thorndyke, 1979). However, those instances of the use of schema theory to account for comprehension cannot be considered fully here for lack of space.

Conclusion

It was my intent in this paper to give the reader unfamiliar with schemata an intuition through which he/she could interpret the increasing number of papers employing these conceptualizations. I have aimed for generality rather than specificity in my account. I have tried to show the many domains to which the concept of a schema has been applied and the heuristic value of thinking about psychological and educational problems in terms of schemata. Although the development of schema-based theories, such as the ones I mention above, is yet in its infancy and these ideas have not yet proved their usefulness, I believe that they offer the most promising leads for those of us interested in the difficult problems posed when we try to apply psychological theories directly to educationally relevant domains.

References

BARCLAY, J.R., BRANSFORD, J.D., FRANKS, J.J., McCARRELL, N.S., & NITSCH, K. Comprehension and semantic flexibility. *Journal of Verbal Learning and Verbal Behavior*, 1974, *13*, 471-481.

BARTLETT, F.C. *Remembering*. Cambridge: Cambridge University Press, 1932.

BOBROW, D.G., & NORMAN, D.A. Some principles of memory schemata. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

BRANSFORD, J.D., BARCLAY, J.R., & FRANKS, J.J. Sentence memory: A constructive versus interpretive approach. *Cognitive Psychology*, 1972, *3*, 193-209.

BRANSFORD, J.D., & JOHNSON, M.K. Considerations of some problems of comprehension. In W.G. Chase (Ed.), *Visual information processing*. New York: Academic Press, 1973.

BRUNER, J.S., & POTTER, M.C. Interference in visual recognition. *Science*, 1964, *144*, 424-425.

CHAFE, W.L. Giveness, contrastiveness, definiteness, subjects, topics, and point of view. In C.N. Li (Ed.), *Subject and topic*. New York: Academic Press, 1976.

FILLMORE, C.J. An alternative to checklist theories of meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistics Society*, 1975, *1*, 123-131.

HEAD, H. *Aphasia and kindred disorders of speech*. New York: Macmillan, 1926.

KANT, E. *Critique of pure reason* (1st ed. 1781; 2nd ed. 1787, N. K. Smith, trans.). London: Macmillan & Co., 1963.

KAPLAN, R.M. A general syntactic processor. In R. Rustin (Ed.), *Natural language processing*. New York: Algorithmics Press, 1973.

MANDLER, J.M., & JOHNSON, N.S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111-151.

MEYER, B.J.F. *The organization of prose and its effects on memory*. Amsterdam: N[illegible]h Holland, 1975.

MINSKY, M. A framework for representing knowledge. In P.H. Winston (Ed.), *The psychology of computer vision*. New York: McGraw-Hill, 1975.

MOORE, J., & NEWELL, A. How can MERLIN understand? In L.W. Gregg (Ed.), *Knowledge and cognition*. Potomac, Md.: Erlbaum, 1973.

NORMAN, D.A., RUMELHART, D.E.; & THE LNR RESEARCH GROUP. *Explorations in cognition*. San Francisco: Freeman, 1975.

PALMER, S.E. Visual perception and world knowledge: Notes on a model of sensory-cognitive interaction. In D.A. Norman, D.E. Rumelhart, & and the LNR Research Group, *Explorations in cognition*. San Francisco: Freeman, 1975.

RUMELHART, D.E. Notes on a schema for stories. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

RUMELHART, D.E. Understanding and summarizing brief stories. In D. LaBerge & S.J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

RUMELHART, D.E., & ORTONY, A. The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Erlbaum, 1977.

SCHANK, R.C., & ABELSON, R.P. Scripts, plans, and knowledge. *Advance papers of the Fourth International Joint Conference on Artificial Intelligence*. Tbilisi, Georgia, USSR, 1975, pp. 151-157.

SPIRO, R.J. Inferential reconstruction in memory for connected discourse. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Erlbaum, 1977.

THORNDYKE, P.W. Knowledge acquisition from newspaper stories. *Discourse Processes*, 1979, *2*, 95-112.

TULVING, E. & THOMSON, D.M. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 1973, *80*, 352-373.

WILLIAMS, M.D. Some observations on the process of retrieval from very long term memory (Doctoral dissertation, University of California, San Diego, 1977). *Dissertation Abstracts International*, 1977.

WINOGRAD, T. Frame representations and the declarative-procedural controversy. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

Comprehension of Text Structures

P. David Pearson
University of Illinois at Urbana-Champaign
Kaybeth Camperell
University of Wisconsin

In this paper, we will review theoretical and empirical developments in the comprehension of text structure over the past twenty years. Following that review, we will offer some suggestions about what this area of scholarship has to say about educational practice. The suggestions will be of two types: a) tentative suggestions about educational practices that educational publishers and/or teachers ought to consider in preparing texts for students and lesson plans to help them cope with variations in text structure, and b) suggestions to educational researchers concerning classroom research which seem reasonable in the light of basic research about text structure influences.

What Would It Mean to Find that Text Structure Influences Comprehension?

Perhaps a good starting point for a review concerned with the influence of text structure on prose comprehension is to ask of what consequence any conceivable findings might be. For example, suppose our review were to demonstrate that 90 percent of the variation in students' comprehension of prose materials was due to the influence of variation in text structure. Suppose we could demonstrate that by holding content (the ideas, concepts, and relations among concepts) of a passage constant and altering the surface structure in which the ideas are communicated we could move a student from 25 percent comprehension of the passage to 75 percent comprehension. What would we recommend? Clearly, we would immediately inform the publishing industry that we had made a breakthrough in communication technology and write manuals on how to communicate ideas effectively in prose.

Suppose, alternatively that we found that variation in text structure had virtually no effect on comprehension. Suppose we found, in the hypothetical experiment above, that such alterations yielded a modest 5 percent instead of a 50 percent gain in comprehension of passages. We would probably drop our heads a little and recommend that our colleagues look elsewhere for any answers to the question of how to improve our communication efficiency—look at the nature of the concepts themselves and relations among them, perhaps.

Suppose that we found that variations in text structure made a big difference for young students but that the differences between various levels of complexity decreased as a function of age. We would likely advise our publisher to avoid certain grammatical structures or text organization patterns until some optimal age level at which, presumably, students have gathered enough linguistic experience to handle their complexity.

Suppose that we found, along with this hypothetical developmental effect, that we could overcome the deleterious effect of certain text structures on younger students by offering them direct instruction in dealing with those structures. We would advise teachers that if they are going to present young students with text embodying those structures, they will have to teach students how to handle them.

Findings like the hypothetical cases described above, while not quite so dramatic as our make-believe examples, have emerged from time to time over the past twenty or thirty years, sometimes but not always accompanied by recommendations like those we have suggested. That text structure influences comprehension, therefore, is not really at issue; what is at issue is the precise way in which the influence is exerted, why the influence exists, and what the influences have to say about practical matters of teaching and writing instructional materials.

What Counts as Text Structure

Before we can review the literature on the influence of text structure, we need to define what we mean by variations in text structure. We will approach our definition through examples. Consider examples (1-4) below.

1. The lad rode the steed to victory.
2. The steed was ridden to victory by the lad.
3. The young man rode the horse to a first place finish.
4. The horse was ridden to a first place finish by the young man.

Each of the four sentences exhibits a different surface form for the same underlying idea, yet there are only two different grammatical structures represented, the active and the passive. Hence the difference between (1) and (2) or between (3) and (4) can be regarded as a variation in text structure. However the differences between (1) and (3) or (2) and (4) are better characterized as lexical, or possibly semantic, variations. In the language of tranformational generative grammar, we would say that (1) and (2) have the identical deep structure (underlying meaning), as do (3) and (4).[1] By intuition we would probably agree that there is only a slight difference in the two deep structures attributable to connotative differences in the meanings of lad-young man, victory-first place finish, and steed-horse. But the basic point is that we will regard alterations in the grammatical structure of sentences, which do not alter any semantic meanings or relations, as examples of text structure manipulations.

Now consider examples (5-8).

5. Henry lost the quarterback job because his arm gave out.

---

[1] Many theorists would argue that the underlying meaning of the active/passive pairs is not identical. Clark and Clark (1977) using a topic-comment formulation, would argue that the focus differs from active to passive: In 1), what is emphasized is some new information about the lad, i.e., that he rode the steed to victory, whereas in 2) the emphasis centers on new information about the horse, i.e., that it was ridden to victory by the lad. We concur. If we adopt a strict equivalence-in-meaning criterion, then the concept of paraphrase (multiple surface structure representations for a single deep structure) cannot exist. Nonetheless, we would probably find that 95 percent of the population would agree that the same basic notions are being communicated in active/passive pairs.

6. Because his arm gave out, Henry lost the quarterback job.

7. Henry lost the quarterback job. His arm gave out.

8. Henry's arm gave out. He lost the quarterback job.

The difference between (5) and (6) is like the difference between (1) and (2) above—simple grammatical transformation, in this case preposing a subordinate clause. The difference between (5) and (7) is not as simple. It is debatable whether or not the causal relation between the two clauses is preserved in (7), unless we resort to the Gricean principle of cooperation between author and reader (Grice, 1975) which posits that no author would arrange the two sentences in (7) adjacent to one another unless he was inviting the reader to infer that the one explained the other. Notice that the invitation to make the causal inference is even stronger in (8), presumably because of the covariation between causal and sequential ordering. Whether variations like those between (5) and (7) qualify as grammatical variation is not clear. But to us they definitely qualify as variation in some aspect of text structure. In certain systems of text analysis they would be regarded as alterations in the rhetorical structure (e.g., Meyer, 1975), cohesion structure (Halliday & Hasan, 1976), or logical structure (e.g., Frederiksen, 1975) of the discourse. Such variations are abundant in naturally occurring discourse, as exemplified in (9-11).

9a. If you want to be a Badger, then come along with me.
9b. Do you want to be a Badger? Come along with me.
10a. After Matthew ate lasagne, he bought a new TV.
10b. Matthew ate lasagne. (Then) he bought a new TV.
11a. Although Susan ran as fast as she could, she lost the race.
11b. Susan ran as fast as she could. But she lost the race.
11c. Susan ran as fast as she could. Alas! She lost the race.

Moving from smaller to larger units of discourse, other kinds of structural variation enter the picture. For example, the structural difference between (12) and (13) has been characterized as a staging variation (Grimes, 1975).

12. Robins build nests in trees. Pheasants build nests in bushes. Eagles build nests in rocks. Birds build nests in a variety of places.

13. Birds build nests in a variety of places. Robins build nests in trees. Pheasants build nests in bushes. Eagles build nests in rocks.

In this case, the meaning of the two texts is similar if not identical; however the position of the rule and its examples is reversed. Notice that this type of variation is a paragraph analogue of positional transformations (active/passive or preposing clauses) at the sentence level.

Thus far we have considered variations in text structure that have only minor influence on the semantic meaning of a text. Furthermore, the structural variations considered occur between or among rather small units of discourse—sentence components or sentences. Such analyses can be regarded as examinations of the microstructure of text.

Other, and particularly more recent, conceptualizations of text structure have ignored the perspective of examining structural variations that preserve meaning in favor of a perspective that examines the hierarchical aspects of text structure. Such schemes usually begin with a parsing of an entire text using either a case grammar (e.g., Meyer, 1975) or a propositional (e.g., Frederiksen, 1975; Kintsch, 1974) scheme to identify relations within and between sentences. Then the entire text is analyzed into a hierarchical structure. Ideas (usually in the form of propositions—basically a clause with an active or stative verb) are scaled according to their structural importance within the hierarchy. For expository texts, importance translates roughly into how "main" or superordinate the idea is. For narrative texts, importance means centrality to the story. Thus, characters, goals and settings are high in the hierarchy while particular episodes or motivations may be fairly low.

Implicit if not explicit in such analyses are two expectations: first, that height in the hierarchy will somehow predict and/or explain the comprehensibility or memorability of particular text segments, and second that surface structures that violate canonical structure will decimate comprehension and recall. Such schemes can be regarded as examinations of the macrostructure of text.

With these two aspects of text structure—microstructure and macrostructure—we have defined the scope of our investigation. Our next step is to examine the empirical studies that have been conducted to evaluate the importance of text structure in comprehension.

Microstructure 1: The Primacy of the Sentence

To psychologists academically reared in the verbal learning tradition of the forties and fifties, the revolution incited by Noam Chomsky's penetrating reviews of behavioral views of language processing (1959) and his alternative views proposed in *Syntactic structures* (1957) and *Aspects of the theory of syntax* (1965), must have seemed a bold departure from the conventional wisdom. The very notion that one could study units of discourse as large and complex as a sentence was revolutionary to researchers more comfortable with lists of nonsense trigrams or quingrams or associations among single words.

Nonetheless, Chomsky's views widened research possibilities for students of verbal behavior. Beginning with the work of Miller and his associates (e.g., Miller, 1962; Miller & Isard, 1963) several researchers in the mid-sixties conducted studies of sentence comprehension. The most common finding among such studies (e.g., Gough, 1965; Slobin, 1966) was that the transformational distance between the underlying meaning of a sentence (its deep structure) and its phonetic realization in speech or graphic realization in writing (its surface structure) was an accurate predictor of the speed or difficulty subjects experience in processing the sentence. In other words, performance could be predicted by variation in the grammar itself. Hence kernel sentences (simple active declaratives) were understood more rapidly than passives, interrogatives, or negatives, which presumably required more cognitive energy to process because more transformations had to be traversed in traveling from surface to deep structure. Findings such as these led to a derivational theory of complexity, i.e., that the derivation of a sentence's surface structure from its underlying deep structure predicted its processing difficulty.

Such studies were appealing to reading educators concerned specifically with reading comprehension. First, they provided needed methodological tools. Finally, there was a way of operationalizing sentence complexity. The notion of a transformation provided a countable index of complexity. In fact, Fagan (1971) and Pearson (1974-75) used a derivational theory of complexity to generate and scale materials used to assess children's comprehension of sentence structure.[2]

Second, these studies corroborated what was known from (or at least implied by) the thirty year old history of readability research: that long complex sentences were associated with passages that rated high in readability and low in comprehensibility.

Unfortunately, the derivational theory of complexity lived only a short life. It was attacked on two different fronts, both as a linguistic theory and as a psychological theory.

The work of Fillmore (1968) on case grammar and generative semanticist framework of linguists like Lakoff (1971) called into question the transformationalists' preoccupation with syntactic relations at the expense of semantic relations.

In psychology, studies such as those conducted by Bransford and Franks (1971) and Sachs (1967) offered data contradictory to the derivational theory of complexity. Implicit in the theory is an assumption that comprehension occurs by analyzing a sentence into its basic constituents (that is how you get from surface to deep structure). Bransford and Franks' evidence suggested that comprehension was better characterized by synthesizing constituents into some semantically integrated chunk. Sachs' data indicated that memory for any aspects of sentence structure faded quite rapidly, while memory for the semantic "gist" of a sentence was remarkably stable. Working with children, Pearson (1974-75) obtained results corroborating with work of Bransford and Franks and Sachs.

---

[2] An almost incidental but nonetheless important methodological tool was the question transformation. Bormuth (1966, 1969, 1971), Pearson (1974-1975), and Bormuth, Manning, Carr, and Pearson (1971) and many researchers since then have used the question transformation as an objective device for generating literal comprehension question probes from text.

In some ways, however, the issue was soon to become a moot point because somewhere in the early seventies researchers turned their attention away from the sentence as a unit of linguistic analysis in favor of stories, passages, and expositions—with a concomitant emphasis on macrostructure rather than microstructure. Later we will examine that line of research; first, however, we must add two pieces to the microstructure puzzle.

Microstructure 2: Linguistic Connectives

The small but interesting body of research dealing with linguistic connectives speaks incidentally to issues of structural variation. This stems from the fact that when a connective is used in a sentence, it often has the effect of increasing the grammatical complexity of the sentence; connectives are involved in the formation of compound sentences and subordinate clauses such as those beginning with *because, although, before,* etc. In a sense, an examination of linguistic connectives is a sensible bridge from the studies looking at the sentence as a unit of analysis to those (in a later section) which emphasize the larger organizational patterns of text. Linguistic connectives usually establish or cue logical relations among propositions or sentences.

Walmsley (1977) defines linguistic connectives as follows:

> A linguistic connective (or logical or language connective—the terms appear to be used according to the orientation of a writer's discipline) may be defined as a "co-ordinating, qualifying or adverbial conjunction used to link a simple proposition with another idea (either a proposition or a concept) to form a complex proposition." Alternatively, it may be defined as a syntactic structure signalling underlying logico-semantic relations (see Olds, 1968). Connectives may link propositions within or between sentences; they may comprise a single word (e.g., *and*), or a phrase (e.g., *in addition to*). (p. 319)

Some researchers have examined the developmental changes that occur across ages in children's understanding of connectives and the relations between the propositions they link (e.g., Beilin & Lust, 1975; Neimark, 1970; Neimark & Slotmick, 1970; Paris, 1973). Not surprisingly, children's understanding improves with age; however, the research in this tradition, because of the nature of the task and isolated (not contextually embedded) stimuli, offers little advice concerning their role in reading larger units of discourse.

Robertson (1968) conducted one of the few educationally oriented studies. Her examination of basal readers used in the intermediate grades revealed that about one in three sentences in her sample employed some sort of connective. Student comprehension of connectives increased from grade four through grade six and was related to listening and reading ability.

Katz and Brent (1968) found that both first and sixth grade children preferred descriptions of causal relations that were made explicit by the use of a connective. This is consistent with the findings of Pearson (1974-1975) who reported that fourth grade students, given a choice as to the surface form in which a causal relation could be stated, preferred to have the relation stated in a grammatically complex subordinated form which included specific cues (*because, so*) denoting causal relationships. Pearson speculated that connectives and complexity (they go together) provided "... a more *unified* conception of the causal relation" (Pearson, p. 174) and that it is the function of connectives to make the causal relationship more explicit. These speculations were strengthened by the findings of a follow-up study in which students were asked to read individual sentences in which a causal relationship was either made explicit by inclusion of a causal connective or left implicit by omitting the connective. Results showed that in almost two-thirds of the cases in which subjects were asked to read sentences containing an unmarked (i.e., implicit) causal relationship, a connective *was included in recall*, thus unifying the relationship and making it explicit. Furthermore, if a sentence was not recalled in a cued, unified form, there was a 50 percent chance that it would not be recalled at all. These findings suggest that connectives have a strong effect on the salience of causal relationships expressed in sentences and may serve to facilitate the integration of ideas in memory.

Finally, Marshall and Glock (1978-1979) found that explicitly stated logical (i.e., causal and relational) structures facilitated the recall of propositional content or discourse for "not-so-fluent" (community college) readers. Recall for "truly-fluent" (college) readers for the same passages was more complete than for community college students and was not affected by the presence or absence of explicitly stated relationships. Structure of recall for good readers reflected a greater degree of differentiation among elements of the underlying structure of the text than did the recall of poorer readers who focused primarily on content. Marshall suggests that these differences are due to the fact that good readers have more well established schemata that can be used to interpret and store the meaning of discourse whereas poorer readers have less complete structure and, therefore, must depend to a greater extent on information explicitly encoded in the surface structure of text.

Microstructure 3: Sentence-Combining

Perhaps the most obvious attempt to determine the influence of direct instruction in the microstructure of text on comprehension has been in the tradition of sentence-combining instruction. Beginning with the observation that attempts to teach formal grammar have little positive effect on students' writing ability (Braddock, Lloyd-Jones, & Schoer, 1963; Mellon, 1969), researchers originally looked to sentence-combining as a way of influencing syntactic maturity in writing (Combs, 1975; Mellon, 1969; O'Hare, 1973). More recently, however, researchers have attempted to determine the effects of sentence-combining training on reading comprehension (Combs, 1975; Fisher, 1973; Hughes, 1975; Hunt & O'Donnell, 1970; Straw, 1979).

Basically, sentence-combining activities require students to integrate into a single sentence information expressed in two or more sentences as in (14) and (15).

14. The boy hit the ball. The boy was tall. The ball was small. He hit it through a window.

15. The tall boy hit the small ball through a window.

The rationale for believing that such instruction could alter writers' syntactic maturity seems obvious; the reason for suspecting a concomitant influence on reading comprehension stems from a view of the language arts that what influences growth in one language capacity will influence growth in another.

While positive effects have been reported on some limited measures of reading comprehension (Combs, 1975; Fisher, 1973; Hughes, 1975), only a study by Straw (1979) has looked at the effects of sentence-combining as they transfer to listening, reading, and writing. Straw found that a sentence-combining training condition affected growth in all three language capacities. However, growth in reading comprehension was limited to an investigator constructed cloze test; it did not affect growth on a standardized test. Interestingly, a complementary sentence-reduction task affected growth in reading comprehension (to a lesser degree than did sentence-combining) but not growth on the writing and listening measures. For purposes of our review, Straw's effects are noteworthy even though the treatment effects do not transfer to a standardized test. There is no good reason to believe that a typical standardized test will be sensitive to such instructional treatments. His results do suggest that attention to microstructure, specifically allowing students to actively manipulate it, pays at least short range dividends in comprehension growth.

These results seem compatible with those in the review of linguistic connectives. Note that linguistic connectives often serve the function of combining ideas that could be expressed in separate sentences. Ironically, then, these two areas of research suggest, in contrast to the earlier work in transformational grammar, that attention to cohesion rather than atomization of sentence elements pays greater dividends.

Macrostructure 1: Narratives

The main purpose of this section is to review research evaluating the influence of the overall structure of narratives on students' comprehension and recall of information presented in texts. Several writers (e.g., Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1977; Thorndyke, 1977) have developed formalisms for analyzing the relations among propositions in stories. Propositions can be related in two ways: by their relative position within the hierarchy of a story and by their rhetorical function.

Like phrase structure grammars applied to sentences (e.g., Chomsky, 1957, 1965), in which rewrite rules dictated a sentence's decomposition (e.g., Sentence ==>Noun Phrase + Verb Phrase, Noun Phrase ==>Determiner + Noun + [Sentence], Verb Phrase ==>Verb + [Sentence]), so story schemata or story grammars specify a set of rewrite rules for decomposing the relations among propositions in a story. Thus a story can be rewritten as STORY ==>SETTING + THEME + PLOT + RESOLUTION; setting can be rewritten as SETTING==>CHARACTERS + LOCATION + TIME, etc. When all the rewrite rules have been applied to a story, what results is an inverted tree diagram for a story, which looks quite similar to a phrase structure parsing of a sentence, except that the basic units are sentences or propositions rather than words.

In essence, this tree structure creates a hierarchy. What appear at the top of the hierarchy are the setting of the story (including characters, location, etc.), the basic theme, a few of the key episodes in the plot, and a resolution of the problem that motivated the characters to whatever actions they undertook to begin with. At lower levels in the hierarchy will be subplots. For example, suppose a character needed to get a car to drive to a beach so that he could dig clams for an important dinner. The activities in the story that were associated with getting the car would appear lower in the hierarchy because they were instrumental in *allowing* a higher level event (getting to the beach) to occur. Further suppose that in order to rent a car, the character had to phone several friends to borrow money. Those events would appear at an even lower level. Such hierarchical relations exist among propositions throughout the story; often the implied link between a higher and an immediately lower level event is causation or enablement (a very weak sister to causation—A allowed or enabled B to occur but did not really compel B to occur).

In addition, some grammars have established intracategory connectors to allow for explicit logical connection between events or states at the same level in the hierarchy. Stein and Glenn (1977), for example, include AND, THEN, and CAUSE links. Hence rhetorical or logical connection between events and states is carried in two ways: vertically by implied hierarchical relations and horizontally by explicit links in the grammar.

Story grammarians have postulated two possible consequences of story grammars. Assuming that students internalize, through constant exposure to stories of various degrees of well-formedness, something like a schema for stories, then comprehension and recall of stories ought to be influenced by two kinds of variation. First, information in higher level nodes ought to be recalled more frequently than that in lower level nodes because of greater centrality to the basic actions and motivations of the characters. Second, violations in the well-formedness of stories (e.g., the degree to which the order of key events is reversed or scrambled, placing motivations out of synchronization with actions, placing setting information at the end of a story, etc.) ought to decrease comprehension and recall.

The first of these predictions has been emphasized by Rumelhart (1975, 1977). He has established a set of story summarization rules to predict the probability that a proposition will be recalled; basically a proposition is predicted to be recalled if a proposition lower in the hierarchy was recalled. He found that the conditional probability that a proposition would be recalled given that it was predicted to be recalled was .95. Rumelhart also interprets the data from the work of Thorndyke (1977) and Meyer (1975) as supporting his hierarchical hypothesis.

Other researchers (e.g., Mandler & Johnson, 1977; Stein & Nezworski, 1978; Thorndyke, 1977) have emphasized the effects in violation of canonical story structure. Thorndyke (1977) found that story recall was debilitated increasingly by a) moving the theme or goal to the end of the story, b) removing the theme altogether, and c) more or less randomly permuting the sentences in the story. Kintsch, Mandel, and

47

Kozminsky (1977) asked college students to read well- and ill-formed (scrambled paragraphs) stories in time-limited or unlimited conditions. Then the students wrote summaries of the 1400 word stories. In the unlimited reading time condition, there was a 23 percent increase in reading time due to scrambling *but* no differences in writing time, length, or quality of story summaries. However in the limited time condition, better summaries were written for well-formed stories. Kintsch et al. felt that subjects in the unlimited time condition imposed a story structure on the scrambled text at the point of comprehension rather than simply at the point of summarization; hence the difference in reading but not summarizing time. Without time to restructure the ill-formed story, comprehension suffered, resulting in inferior summaries. Stein and Nezworski (1978) found results similar to those of Thorndyke (1977). Well-formed stories elicited better story recall than stories containing slightly disordered, randomly ordered, or unrelated statements. Furthermore, unrelated statements elicited the greatest number of inferences into recall, reflecting subjects' attempts to make sense out of an incoherent text, a finding reminiscent of Bartlett's (1932) early results on cross-cultural intrusions into story recall.

The developmental (cross-age) data collected by Stein and Glenn (1977) and Mandler and Johnson (1977) also support the notion of story schemata. As children grow older they tend to recall increasingly more of the lower level information in the story. Young children tend to recall only a few of the higher level propositions such as a character, an initiating event, and an outcome.

While story schemata have been criticized for their emphasis on prediction rather than explanation and the fact that they predict too many behaviors (Thorndyke & Yecovich, in press), their basic validity as formalizations of what people learn when they learn about how writers put stories together seems to us to be well-founded.

Instructional Research on Story Schemata

We were able to locate only three studies dealing even tangentially with issues of direct instruction about how stories are structured. Bower (1976) had subjects read a biography about a fictitious poet. Then half the subjects read two biographies with similar macrostructures while half read two unrelated texts. When they were asked to recall the original biography, experimental subjects (the three biographies) recalled more of the macrostructure (which was similar in all three) but interconfused details of the second and third with the first. The similarity of the three passages created macrostructure facilitation and detail interference. Thorndyke (1977) found that subjects who read a second story with the same structure as, but different characters from, the first story recalled more second story information than those whose second stories had the same characters as, but a different structure from, the first.

Neither of these studies can be considered instructional in anything but an incidental sense. However, a study by Gordon (1980) speaks directly to story structure instruction. Over a period of eight weeks she trained fifth grade students to apply a simplified story schema to basal reader stories that they read as a part of their normal reading instruction. On a transfer story, these students recalled significantly more, particularly of certain categories of high-level information, than a placebo or an untreated control group. She interpreted the findings as supporting the notion that direct instruction in story schemata provides students with a transferable framework for storing and retrieving textually presented information.

Macrostructure 2: Exposition

Research and theory about the macrostructure of expository text is not quite so abundant as that for narrative. Attempts have been made by Kintsch (1974), Frederiksen (1975), and Meyer (1975) to develop general schemes for representing relations among units of text. Kintsch and Frederiksen give more emphasis to a scheme that could serve as either a model of text structure or the structure of knowledge in memory, Meyer's system is, admittedly, more concerned with representing text per se. Because of space limitations and because it places greater emphasis on text macrostructure, we have chosen to concentrate on Meyer's system, recognizing full well that we can justify our decision only by asserting that we intend our review to represent an example from a class of text structure schemes.

Adhering closely to the theory of Grimes (1975) for connected discourse and to Fillmore's (1968) case grammar, Meyer has developed a text structure system that emphasizes relations among propositions in a text. She has lexical propositions that show the case relations between words within simple sentences and clauses. And she has rhetorical propositions which establish the relations between and among sentences, paragraphs, and longer units of text. Rhetorical predicates are labels used to specify the relationships within these propositions. Rhetorical predicates order the ideas in a text into hierarchical relationships, and they allow Meyer to develop a richer, higher-level organization than either Kintsch or Frederiksen.

Meyer's parsing of a passage looks much like an outline of the passage, except that all the ideas from the passage are included. Top-level discourse structures in the outline are simply the relations that occur in the top third of the diagram. Height in the system is indicated by "leftness" of a proposition in the content outline.

A basic thesis of Meyer's is that height in the hierarchy predicts how well propositions will be comprehended and recalled. She designed an experiment in which a target paragraph was embedded high within the hierarchy of passage 1 but low in passage 2. The serial position of the paragraph was identical across passages. While she found no overall recall differences between the two passages, the target paragraph was recalled better when it was staged higher in the hierarchy. These immediate recall differences increased with a week's delay. Similar differences were noted in the cued recall of the target passage after a week's interval.

Meyer (1977a, 1977b) extended her research to determine whether or not sixth grade students were sensitive to these hierarchical differences in content structure. Meyer predicted that students classified as low in ability would recall more information from low levels of the content structure than from high levels in the content structure.

Immediately after listening to a short article about parakeets, students answered fifteen main idea and fifteen detail questions about the article. The main idea questions were derived from idea units high in the structure of the passage, and the detail questions were derived from idea units low in the structure of the passage. Results indicated that all of the students, regardless of ability level, answered more main idea questions than detail questions. Brighter students remembered significantly more information from both levels of the structure than other students, but even low-ability students answered more main idea questions correctly than detail questions. Meyer concluded that children, like adults, remember more information from high levels of the content structure of a text and that a content structure representation can be useful in generating different types of comprehension questions for prose materials. Meyer cautions, however, that the results of this study might not generalize to low-ability students with reading or learning disabilities under reading versus listening conditions.

In order to explore the effects of different types of top-level discourse structures on recall, Meyer and Freedle (1979) had graduate students read articles with identical middle- and low-level structures and content. The passages differed in the way similar introductory information was organized in the top-level of their content structure diagrams. The four types of structures (rhetorical predicates she calls them) compared in the study were: adversative (contrastive pattern), covariance (cause-effect pattern), response (problem-solution pattern), and attributive (list-like pattern). The investigators predicted that information in passages organized with adversative, covariance, and response structures would be remembered better than information from the passage organized with an attributive structure.

Subjects participating in the study were graduate students working on advanced degrees in education. They were divided into four groups, and each group listened to a passage organized by one of the four rhetorical predicates. An immediate free-recall test, a delayed free-recall test, and a delayed short-answer test were administered to all subjects. The short-answer test consisted of questions which tapped memory for information that was identical in each of the four passage conditions. Recall protocols were scored for the number of idea units recalled and for the type of rhetorical structure subjects used to organize their recall protocols. The short-answer test was simply scored for the number of correct answers.

Subjects who listened to passages organized with adversative (constrastive) and covariance (cause-effect) structures remembered significantly more information than subjects who listened to passages organized with attribution (list-like) and response (problem-solution) structures. Moreover, subjects who listened to the adversative passage answered significantly more of the short-answer questions correctly than subjects who listened to the other passages. Subjects who listened to passages with adversative and covariance structures also used these types of relationships to organize their recall protocols.

From these findings, Meyer and Freedle concluded that differences in the type of structure used to organize textual information significantly affected the amount of information graduate students learned and remembered. Adversative and covariance organizations enhanced recall over attribution and response organizations.

Using a schema theory orientation, Meyer and Freedle had predicted that adversative, covariance, and response structures would provide better organization for learning than an attributive, list-like structure. Each of the four types of structure is used in expository texts to let readers know information will be presented about a topic; but adversative, covariance, and response structures ostensibly provide readers with additional schemata to help them understand and remember the information. For example, an adversative structure indicates that the information will be about opposing views; covariance structures indicate that the information will be about causal relations; and response structures indicate that the information will be about problems and solutions. Attributive structures are more loosely organized, however, and do not provide additional information.

The prediction that adversative and covariance structures would facilitate recall was supported. The prediction about the response structure, however, was not confirmed. Meyer and Freedle explain this unexpected finding in terms of social-psychological facts and notions of perspective. The subjects participating in the study were school teachers who may have been offended by the solution in the response passage as it involved firing coaches. Thus, the teachers seemed to reject the schema provided by the author, read the text from their own perspective or personal viewpoint, and thereby processed the text differently than was expected.

The most important finding in this study was that certain types of top-level discourse structures did facilitate recall more than others. Meyer and Freedle interpret the results of this study as showing that the most efficient strategy students can adopt in typical school-learning or lab-learning situations is to identify and use the author's organizational framework to guide and structure their attempts to understand and remember information from textual materials. Students who are familiar with the way texts are typically organized can use that knowledge to comprehend and remember by relating the organizational structure, or schema, of the text to their prior knowledge (stored schemata) about how texts are organized and what to expect from texts organized in certain ways.

Meyer, Brandt, and Bluth (1978) investigated the effects of identifying and using the organizational structure of texts on recall. They predicted that readers who adopted the strategy of identifying the author's organizational structure (the author's schema) would be able to recall more information than students who did not adopt this strategy. Ninth grade students classified as good, average, poor, and "difference" (high vocabulary but low comprehension scores) readers participated in the study. They read and recalled two different expository passages. One passage was organized with a response predicate and the other with an adversative predicate. Thus, the passages differed in their top-level rhetorical structures and, also, in whether or not signaling devices were present in the texts.

Signaling devices, as defined by Meyer (1975), are ways in which authors emphasize aspects of the semantic content or structure of a text. The title of the passage and words such as "in contrast to" were types of signaling used in the adversative passage. An explicit statement of the problem and solution relations as well as signaling words such as "first," "second," etc. were the types of signaling included in the response passage. Meyer et al. predicted that signaling devices would benefit poor and "difference" readers in processing the texts as it was assumed that these readers did not normally use the organizational structure of texts to understand and remember information.

Immediate and delayed free-recall tests were scored for the number of idea units recalled and the degree of similarity between the organization of the recall protocols and that of the original passages. Results indicated that good readers organized their protocols with the same structure as that used in the passages they read and that they recalled significantly more information than students who did not adopt this strategy. This result was obtained with good readers even when signaling devices were not present in the texts they read. The strategy of using the author's "schema" to organize recalled information was a better predictor of recall than either standardized comprehension or vocabulary test scores; multiple regression analyses indicated that use of this strategy accounted for 44 percent of the variance in recall on the immediate-recall test and 68 percent of the variance in recall on the delayed test.

Signaling appeared to facilitate recall of low and average comprehenders on the immediate test but not on the delayed-recall test. On the immediate test, the students classified as poor, average, and "difference" readers who read the response passage with signaling organized their recall protocols with the same pattern of relationships as those in the original passage; they also recalled significantly more information than similarly classified students who read the without-signaling version of the passage. However, signaling had no effect on the recall of students who read the without-signaling and adversative versions of the passage.

In a subsequent study, Bartlett (1978) taught a group of ninth grade students to identify various types of top-level structures common to expository texts and to use the structures to organize their recall protocols. The students were taught how to identify and to use covariance, adversative, attribution, and response structures during a week-long training period. Apparently, students trained to use the strategy of identifying an author's top-level structure were able to recognize these structures in texts significantly better than students who did not receive training; and trained students were able to recall nearly twice as much information after reading than students who did not receive training. Thus, some evidence exists to suggest that students can be taught to identify top-level discourse structures and that such training improves comprehension.

Implications for Reading Practice

It is always somewhat dangerous to leap too boldly across the gap from research, especially basic research, to educational practice. A more cautious approach is to suggest that research findings from laboratory or other basic research settings should be regarded as grist for applied research studies which should be carried out in real school environments before we make any conclusive recommendations for changes in materials or teaching strategies (e.g., Bronfenbrenner, 1976). Nonetheless, we see several areas in which the leap seems so reasonable and inviting that we make it, caution notwithstanding.

Recommendations to People Who Prepare Reading Materials

1. The research on children's comprehension of story structure suggests to us that from the outset of grade one, children ought to be reading stories that are highly predictable in terms of their conformity to canonical story schemata. We recognize that the need to control vocabulary in the earliest of stories makes it difficult for writers to create well-formed stories. Yet we are convinced that it is these young children who need the predictability the most. Consider the case of a first grade student

who is trying to make sense out of the unfamiliar orthography of English writing. The child is already confronted with one source of potential confusion (figuring out what sounds the letters make); to embed that task in a context that can be another source of confusion (stories that violate story schemata) seems to compound the problem.

2. The research on the influence of connectives, structures of cohesion, and sentence combining activities suggests that complexity may sometimes add to rather than always detract from the likelihood that comprehension will occur. Cohesive forms of statements appear to make explicit what is otherwise left to children's inferential powers. Textbook writers need to be aware of this fact. Above all they should not be led to the false conclusion that writing becomes more readable when complex sentences are chopped in half (even though such a practice will reduce a passage's readability scores).

Recommendations to Educators (and Writers of Instructional Practice Materials)

3. The salutary on sentence-combining training has been replicated several times. It seems reasonable to recommend that students be given an opportunity to learn an important fact about the English language: that there is always more than one way to express a given idea. Awareness of this fact also apparently leads to growth in syntactic maturity and listening, particularly if the focus is on creating cohesive statements.

4. If teachers want students to "get the author's message," they are well advised to model for students how to figure out what the author's general framework or structure is and then allow students to practice discovering it on their own. They should be cautioned, however, that not all reading has as its purpose "getting the author's message"; sometimes students need to read to update their own knowledge, in which case they are probably better off working within their own schemata rather than an author's schema (Spiro, 1977).

We close this section with a disclaimer: These are not the only suggestions which could be derived from the research base on comprehension of text structure, only those that seemed most reasonable to use. Also, these may not be the most important implications to be derived for practice from research; it may be that the research based on the structure of knowledge in memory or the interaction between text structure and knowledge structure or the process of learning to monitor one's own comprehension may prove more fruitful for instruction. They simply do not fall under the scope of this review.

Implications for Instructional Research in Reading

1. We need to know more about the point in time when children are able to handle certain complex kinds of syntactic structures. There was a time in the late sixties when the conventional wisdom concerning syntactic development seemed to suggest that, by the age of six, children had mastered nearly all the syntactic structures they would use as adults. Then the work of C. Chomsky (1969), Bormuth, Manning, Carr, and Pearson (1971), Olds (1968), and others pointed out that even by age ten children still had trouble within many structures. Somehow the rush toward semantic and macrostructural concerns in the mid-seventies buried what was an incomplete and fruitful line of research. We still need to finish the job.
2. After the issue of development comes instructions: Are those structures which cause difficulty even for the ten-year-old amenable to direct instruction and systematic practice?
3. The work of Meyer and her associates suggests that good readers are better at following an author's rhetorical plan of organization than are poor readers. The next step is to demonstrate that poor readers who receive direct instruction in deciphering an author's organizational plan improve in their ability to produce greater veridical comprehension and recall of text.
4. In this regard, we need to know more about the relative efficiency of different rhetorical plans of organization (e.g., adversative, covariance, attributional, etc.) in communicating content in various disciplines (science versus history versus geography). It may be that certain plans are uniquely suited to certain types of content.

5. The work of Gordon (1980) should be extended to younger age levels to see if the salutory effects of story schema training will assist even younger students. In this regard, we should mention the exciting but emerging work of a group of researchers in Boston (Rubin, 1980) who are using a story schema framework to help young children get off to a faster start in writing as well as reading stories.

6. Finally let us offer one general suggestion for instructional research derived from constructs emanating from basic research. When we look at research on teaching and learning variables, we have been awestruck by the persistence and ubiquity of two terms: engaged time on task and direct instruction (e.g., Becker, 1977; Berliner, 1975; Rosenshine, 1976). We finally seem to be getting the message that kids learn what they are taught and get to practice. Thus far, the research seems to have shown these effects in more mundane aspects of reading such as word identification. But there is no reason to believe that they wouldn't aid comprehension as well, even though we have evidence that few teachers teach comprehension (Durkin, 1978-1979). In fact, the work of Straw (1978) and Gordon (1980) reported earlier, as well as a recent study by Hansen (1979) seem to provide direct evidence that students learn new strategies for comprehending text when they are taught and practiced systematically. The point is simple: when we identify a variable, including a text structure variable, that looks like it might make a difference in comprehension, we ought to adopt a frontal assault strategy when considering its instructional power—teach about it systematically and make certain students have a chance to practice it. The time for a renaissance of the methodological study is now—now that we have a better idea of what to look at.

A final caution: we don't expect that the products of this new methodological research will be altogether new and surprising. In fact, we expect that many will elicit reactions of "reinventing the wheel," or "that's just common sense." Such reactions will please us. Common sense is all too common and all too sensible to be overturned by a single line of research. But no real value in the new research will be the contextual and

58

theoretical base from which it emanates. Hence we will be in a better position to answer the question, Why did it work?

References

BARTLETT, B.J. *Top-level structure as an organizational strategy for recall of classroom text.* Unpublished doctoral dissertation, Arizona State University, 1978.

BARTLETT, F.C. *Remembering: A study in experimental and social psychology.* Cambridge, Eng.: Cambridge University Press, 1932.

BECKER, W.C. Teaching reading and language to the disadvantaged—What we have learned from field research. *Harvard Educational Review*, 1977, *47*, 518-543.

BEILIN, H., & LUST, B., et al. *Studies in the cognitive basis of language development.* New York: Academic Press, 1975.

BERLINER, D.C. *Impediments to the study of teacher effectiveness* (Tech. Rep. No. 75-11-3). San Francisco, Calif.: Far West Laboratory for Educational Research and Development, November, 1975.

BORMUTH, J.R. Readability: A new approach. *Reading Research Quarterly*, 1966, *1*, 79-132.

BORMUTH, J.R, MANNING, J.C., CARR, W., & PEARSON, P.D. Children's comprehension of between- and within-sentence syntactic structures. *Journal of Educational Psychology*, 1971, *61*, 349-357.

BOWER, G.H. Experiments on story understanding and recall. *Quarterly Journal of Experimental Psychology*, 1976, *28*, 511-534.

BRADDOCK, R., LLOYD-JONES, R., & SCHOER, L. *Research in written composition.* Urbana, Ill.: National Council of Teachers of English, 1963.

BRANSFORD, J.D., & FRANKS, J.J. The abstraction of linguistic ideas. *Cognitive Psychology*, 1971, *2*, 331-350.

BRONFENBRENNER, U. The experimental ecology of education. *Educational Researcher*, 1976, *5*, 5-15.

CHOMSKY, C. *The acquisition of syntax in children from 5 to 10.* Cambridge, Mass.: MIT Press, 1969.

CHOMSKY, N. *Syntactic structures.* The Hague: Mouton, 1957.

CHOMSKY, N. A review of *Verbal Behavior*, by B.F. Skinner. *Language*, 1959, *35*, 26-58.

CHOMSKY, N. *Aspects of the theory of syntax.* Cambridge, Mass.: MIT Press, 1965.

CLARK, H.H., & CLARK, E. *Psychology and Language.* New York: Harcourt Brace Jovanovich, 1977.

COMBS, W.E. *Some further effects and implications of sentence-combining exercises for the secondary language arts curriculum.* Unpublished doctoral dissertation, University of Minnesota, 1975.

DURKIN, D. What classroom observations reveal about reading comprehension instruction. *Reading Research Quarterly, 14,* 1978-1979, 481-533.

FAGAN, W.T. Transformations and comprehension. *The Reading Teacher,* 1971, *25,* 169-172.

FILLMORE, C.J. The case for case. In E. Bach & R. Harms (Eds.), *Universals in linguistic theory.* New York: Holt, Rinehart & Winston, 1968.

FISHER, K.D. *An investigation to determine if selected exercises in sentence-combining can improve reading and writing.* Unpublished doctoral dissertation, Indiana University, 1973.

FODOR, J.A., & GARRETT, M.F. Some syntactic determinants of sentential complexity, *Perception and Psychophysics,* 1967, *2,* 289-296.

FREDERIKSEN, C.H. Representing logical and semantic structure of knowledge acquired from discourse. *Cognitive Psychology,* 1975, *7,* 371-458.

GORDON, C.J. *The effects of instruction in metacomprehension and inferencing on children's comprehension abilities.* Unpublished doctoral dissertation, University of Minnesota, 1980.

GOUGH, P.B. Grammatical transformations and speed of understanding. *Journal of Verbal Learning and Verbal Behavior,* 1965, *4,* 107-111.

GRICE, H.P. Logic and conversation. In P. Cole & J.L. Morgan (Eds.), *Syntax and semantics, 3, Speech acts.* New York: Seminar Press, 1975, 41-58.

GRIMES, J.E. *The thread of discourse.* The Hague: Mouton, 1975.

HALLIDAY, M.A.K., & HASAN, R. *Cohesion in English.* London: Longman, 1976.

HANSEN, J.A. *The effects of two intervention techniques on the inferential ability of second grade readers.* Unpublished doctoral dissertation, University of Minnesota, 1979.

HUGHES, T.O. *Sentence-combining: A means of increasing reading comprehension.* Kalamazoo: Western Michigan University, Department of English, 1975.

HUNT, K.W., & O'DONNELL, R. *An elementary school curriculum to develop better writing skills.* U.S. Office of Education Grant No. 4-9-08-903-0042-010. Tallahassee: Florida State University, 1970.

KATZ, E., & BRENT, S. Understanding connections. *Journal of Verbal*

Learning and Verbal Behavior, 1968, *7*, 501-509.

KINTSCH, W. *The representation of meaning in memory*. Hillsdale, N.J.: Erlbaum, 1974.

KINTSCH, W., MANDELL, T.S., & KOZMINSKY, E. Summarizing scrambled stories. *Memory and Cognition*, 1977, *5*, 547-552.

LAKOFF, R. If's, and's, and but's about conjunction. In C.J. Fillmore & D.T. Langendoen (Eds.), *Studies in linguistic semantics*. New York: Holt, Rinehart & Winston, 1971.

MANDLER, J.M., & JOHNSON, N.S. Rememberance of things parsed. Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111-151.

MARSHALL, N. *Comprehension of connected discourse: A study into the relationships between the structure of text and information recalled.* Unpublished paper, Arizona State University, 1976.

MARSHALL, N., & GLOCK, M. Comprehension of connected discourse: A study into the relationships between the structure of text and information recalled. *Reading Research Quarterly*, 1978-1979, *16*, 10-56.

MELLON, J. *Transformational sentence combining: A method for enhancing the development of syntactic fluence in English composition* (NCTE Research Rep. No. 10). Urbana, Ill.: National Council of Teachers of English, 1969.

MEYER, B.J.F. *The organization of prose and its effects on memory*. Amsterdam: North-Holland Publishing, 1975.

MEYER, B.J.F. The structure of prose: Effects on learning and memory and implications for educational practice. In R.C. Anderson, R. Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Erlbaum, 1977. (a)

MEYER, B.J.F. What is remembered from prose: A function of passage structure. In R. O. Freedle (Ed.), *Discourse production and comprehension* (Vol. 1). Norwood, N.J.: Ablex, 1977. (b)

MEYER, B.J.F. *Structure of prose: Implications for teachers of reading* (Research Report No. 3). Tempe: Arizona State University, Prose Learning Series, Spring 1979.

MEYER, B.J.F. & FREEDLE, R. *The effects of different discourse types on recall*. Princeton, N.J.: Educational Testing Service, 1979.

MEYER, B.J.F., BRANDT, D.M., & BLUTH, G.J. *Use of author's schema: Key to ninth graders' comprehension*. Paper presented at the meeting of the American Educational Research Association, Toronto, March 1978.

MILLER, G.A. Some psychological studies of grammar. *American Psychologist*, 1962, *17*, 748-762.

MILLER, G.A., & ISARD, S. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*, 1963, *2*, 217-228.

NEIMARK, E.D. Development of comprehension of logical connectives:

Understanding of "or." *Psychonomic Science*, 1970, *21* (4), 217-219.

NEIMARK, E.D., & SLOTNICK, N.S. Development of the understanding of logical connectives. *Journal of Educational Psychology*, 1970, *61* (6), 451-460.

O'HARE, F. *Sentence-combining: Improving student writing without formal grammar instruction* (NCTE Research Rep. No. 15). Urbana, Ill.: National Council of Teachers of English, 1973.

OLDS, H.F. *An experimental study of syntactical factors influencing children's comprehension of certain complex relationships* (Rep. No. 4). Cambridge, Mass.: Harvard University, Center for Research and Development on Educational Differences, 1968.

PARIS, S.G. Semantic and constructive aspects of sentence memory in children. *Developmental Psychology*, 1973, *9* (1), 109-113.

PEARSON, P.D. The effects of grammatical complexity on children's comprehension, recall, and conception of certain semantic relations. *Reading Research Quarterly*, 1974-1975, *10*, 155-192.

REDER, L.M. *Comprehension and retention of prose: A literature review* (Tech. Rep. No. 108). Urbana: University of Illinois, Center for the Study of Reading, 1978.

ROBERTSON, J. Pupil understanding of connectives in reading. *Reading Research Quarterly*, 1968, *3*, 387-417.

ROSENSHINE, B. Classroom instruction. In N.L. Gage (Ed.), *The psychology of teaching methods* (7th NSSE Yearbook). Chicago: University of Chicago Press, 1976.

RUBIN, A. *Making stories, making sense* (Reading Ed. Rep. No. 00). Urbana: University of Illinois, Center for the Study of Reading, 1980.

RUMELHART, D.E. Notes on a schema for stories. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science*. New York: Academic Press, 1975.

RUMELHART, D.E. Understanding and summarizing brief stories. In D. LaBerge & S.J. Samuels (Eds.), *Basic processes in reading: Perception and comprehension*. Hillsdale, N.J.: Erlbaum, 1977.

SACHS, J.S. Recognition memory for syntactic and semantic aspects of connected discourse. *Perception and Psychophysics*, 1967, *2*, 437-442.

SLOBIN, D.I. Grammatical transformations and sentence comprehension in childhood and adulthood. *Journal of Verbal Learning and Verbal Behavior*, 1966, *5*, 219-227.

SPIRO, R. Remembering information from text: The "state of schema" approach. In R.C. Anderson, R. Spiro, & W. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, N.J.: Erlbaum, 1977.

STEIN, N.L., & GLENN, C.G. *A developmental study of children's construction of stories.* Paper presented at the SRCD meetings, New Orleans, March 17-20, 1977.

STEIN, N.L., & NEZWORSKI, T. *The effects of organization and instructional set on story memory* (Tech. Rep. No. 129). Urbana: University of Illinois, Center for the Study of Reading, 1978.

STRAW, S.B. *An investigation of the effects of sentence-combining and sentence-reduction instruction on measures of syntactic fluency, reading comprehension, and listening comprehension of fourth-grade students.* Unpublished doctoral dissertation, University of Minnesota, 1978.

STRAW, S.B. Measuring the effect of sentence-combining instruction on reading comprehension. In D. Daiker, A. Kerek, & M. Morenberg (Eds.), *Sentence combining and the teaching of English.* Akron, Ohio: L. & S. Books, 1979.

THORNDYKE, P.W. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, *9*, 77-110.

THORNDYKE, P., & YECOVICH, F. A critique of schemata as a theory of human story memory. *Poetics*, in press.

WALMSLEY, S. Children's understanding of linguistic connectives: A review of selected literature and implication for reading research. In P.D. Pearson (Ed.), *Reading: Theory, research, and practice.* 26th Yearbook of the National Reading Conference. Clemson, S.C.: The National Reading Conference, 1977.

On the Making of Inferences During Reading and Their Assessment

Thomas Trabasso
University of Chicago

In this paper, I would like to indicate how the making of inferences plays a role in the comprehension of narratives. In so doing, seven questions are posed and answers to each question are discussed. We ask, first, for *a definition of what is meant by inferences.* Then we explore *what functions are performed by inferences.* This is followed by a consideration of *what is required to make inferences* and *what processes are involved in making inferences.* Next, we discuss *what kinds of inferences there are* and take up the relationship between the kinds of inferences and *how one can assess a child's ability to make inferences.* Finally, we end on the practical and educational question of *whether one can promote comprehension through the asking of inferential questions during reading.*

What Are Inferences?

What does a reader do when he makes an inference? From our perspective, he does one of two things: he either finds semantic and/or logical relations between propositions or events which are expressed in the narrative or he fills in missing information which is necessary to making such connections between events. The first kind of inferencing has been called "text

---

The writing of this paper was supported by National Institute of Education grant NIE-G-77-0018 to T. Trabasso and, in addition, under contract No. US-NIE-C-400-76-0016.

connecting" and the second, "slot-filling" (Warren, Nicholas, & Trabasso, 1979). These descriptive terms for making inferences come from a recent theory of natural language understanding and memory (Schank, 1975). According to this view, the process of understanding is largely one of translating a series of sentences into a *causal* chain of underlying conceptualizations. Each sentence consists of one or more conceptualizations which must be derived from the surface structure expressed in the text. This process involves linguistic and world knowledge about individual word meanings and relations within a sentence. The process of creating the causal chain, however, involves inference generation: The reader is assumed to read a story to generate the causal chain and the memory representation, and to encode events that are explicit along with those that are inferred. This representation in memory is then used to perform a variety of operations such as retelling or recalling the story, summarizing the story, detecting the main ideas, deciding which events occurred in which temporal order, answering probe questions as to causes, consequences or facts, paraphrasing events, and giving different points of view of the narrative.

All of the above activities may be recognized as either related to comprehension or susceptible to being captured in comprehension tasks (cf. Pearson & Johnson, 1978). The important contribution of Schank (1975) is the stress on the initial understanding by the reader and on the question of what representation of the story is constructed as a result of this understanding at the time of reading. If the reader (or listener) should fail to construct the relations between events, explicit or inferred, then the subsequent activities would not be possible, the reader having no memorial basis for performing them: Representation results from and requires an initial understanding of sentences and their relations and, in turn, precedes all other forms of comprehension.

The assumption here is that the representation or understanding of a story is essentially a chronology of alternating events and states with causal links. This idea is hardly novel. In fact, Dewey (1963) seemed to have had a similar notion in mind when, in 1933, he wrote on "meaning."

> To grasp the meaning of a thing, an event or a situation is to see it in its relations to other things; to note how it operates or functions, what consequences follow from it; what causes it, what uses it can be put to (p. 135).

In the above quote, note the emphasis on relations to other things, notably causes and consequences.

Since the making of inferences is a highly automatic and largely unconscious process, it is necessary at the outset to use illustrations, both to demonstrate what is meant by an inference and to make it clear that the making of an inference, while highly automatic, is not a simple or obvious process. This should become clearer when we try to understand what the process is and how it is made to operate.

Consider the following pair of sentences, taken from Bransford and McCarrell (1975):

1. John missed the bus.
2. He knew he would have to walk to school.

Note first that there is no explicit causal connection between (1) and (2). Therefore the reader, when confronted with this pair of sentences, would have to make assumptions about the connections between (1) and (2) in order to understand them. If these sentences occurred in the order (1), then (2), the reader might infer that (1) was the causal antecedent of (2) and provide the connective "so," "and then," "thus," or "as a result," etc. The fact that we automatically assume (1) to be the cause of (2) becomes more apparent when we try to interpret the following sentence:

3. John missed the bus *because* he knew he would have to walk to school.

In (3), the cause/consequence relations of (1) and (2) are now reversed. Presumably, John wanted to walk to school *and so* he may have deliberately missed the bus.

Again, consider two more examples from Bransford and McCarrell (1975):

4. The mirror broke.
5. The child grabbed the broom.

We automatically assume that event (4) is the temporal and causal antecedent of event (5) and we fill in the relation as expressed by connectives such as "so" or "therefore." However, our assumptions about cause and effect are apparent when we encounter event (6) which is contrary to the assumed cause/consequence relation.

6. The mirror broke *because* the child grabbed the broom.

When connectives or relations are not explicitly flagged by syntactic markers in text, then readers infer them based upon temporal sequence and causal knowledge of the world. When connectives are explicitly stated, they are used to guide assumptions about causes or consequences in order to comprehend what we read. It makes a great deal of difference in events (3) and (6) how we interpret John's or the child's motives and responsibilities for actions or consequences.

Nicholas and Trabasso (1979) cite another example which we shall use to illustrate first what inferences are and which ones appear to be necessary to understanding text. Then, in the next section, we shall use the example to illustrate functions of inferences.

Suppose you heard the line:

7. Mary had a little lamb.

What do you think of? Nursery rhymes? Mother Goose? Little girls? Fleecy frolicking lambs? Now, read event (7) in conjunction with each of the following events and note how your interpretation shifts.

8. Its fleece was white as snow.
9. She spilled gravy and mint jelly on her dress.
10. The delivery was a difficult one and afterwards the vet needed a drink.

What assumptions appear necessary to understand event pairs (7) and (8), (7) and (9), and (7) and (10)? In event pair (7) and (8), we use our knowledge to infer that Mary is a character from a well-known nursery rhyme—a little girl who is followed about by her pet lamb. The verb "had" alludes to ownership, and the animal is alive and well.

In (7) and (9), the sheep has not fared so well. Here "Mary" is probably human and female since the pronoun "she" and the noun "dress" allow this inference. "Mary" may also be a child since children are more likely to spill food on themselves. The references to gravy and mint jelly indicate, however, that the lamb is actually a meal, not a pet.

Finally, in (7) and (10), the references to the veterinarian and to a difficult delivery suggest that Mary had given birth to a small lamb and is, herself, a mature, female sheep. The vet is probably an adult, human being whose profession is to tend to sick animals. The drink is likely to be alcoholic and is presumably taken to enable the vet to relax after the difficult delivery of the newborn lamb.

Note the vast range of assumptions and knowledge that is necessary to understand these pairs of events. We need to know about nursery rhymes, ownership, pets, little girls, sheep, food, animal births, veterinarians, and alcohol. This knowledge is used to construct an interpretation of (7) in the light of (8), (9) or (10). Note, also, that (7) is an inherently ambiguous sentence and that events (8), (9) or (10) invoke knowledge about three radically different contexts in order to infer information that is implicit in the message. The activation of the knowledge contained in (8), (9) or (10) appears necessary in order to interpret (7) in each of its various meanings.

What Functions Do Inferences Perform?

Inferences perform a variety of functions and by indicating this diversity through definitions and examples, we may more fully appreciate their complexity.

First, intended meanings of individual words are often ambiguous and must be arrived at inferentially. Thus, one function of inferences is to resolve lexical ambiguity. In the above "Mary" sentences, for example, the word "had" may be interpreted respectively as:

owned or possessed (events 7 and 8)
ate (events 7 and 9)
or gave birth to (events 7 and 10)

"Lamb" may be interpreted, respectively, as:
a living animal (events 7 and 8)
a prepared meal (events 7 and 9)
a newborn sheep (events 7 and 10)

A second function of inferences is to resolve nominal and pronominal references (anaphora). Again, in the above examples:

"Its" refers to the lamb and not Mary in (7) and (8).
"She" refers to Mary and not the lamb in (7) and (9).
"Having lamb" refers to delivery or birth in (7) and (10).

In order to interpret sentences while we read, we need to establish a context. This context is also arrived at inferentially. In the above examples, three contexts or topics are inferred:

nursery rhyme in (7) and (8)
meal in (7) and (9)
birth in (7) and (10)

A related, fourth function is that inferences aid in establishing a larger framework for interpretation. We shall now present three sentences used by Collins, Brown, and Larkin (in press) to illustrate how we construct and reconstruct "models" (frameworks) from given information. When one is processing the sentences given as data for constructing a framework, the procedure is said to be "bottom-up." Once the model is constructed and is used to interpret new information, the processing is said to be "top-down." The initial step, upon reading sentence (11), is "bottom-up," but once the "model" is established, we use it in a "top-down" to guide further interpretation. Some "models" are inappropriate or cannot accommodate the subsequent events and are, hence, abandoned. New "models" must be inferred. So, read and think about your "models" as you progress through events (11), (12) and (13):

11. He plunked down $5.00 at the window.
12. She tried to give him $2.50 but he refused to take it.
13. So when they got inside, she bought him a large bag of popcorn.

In studying (11), Collins et al. (in press) found that subjects interpreted the window as that at a racetrack and the $5.00 a bet. Probably, the verb, "plunking down," led to this interpretation since this term is jargon used by bettors for the act of making bets. However, this "model" undergoes reinterpretation in (12) since the attempt to give back $2.50 and its reaction are incongruous with the amounts normally bet at racetracks and with what appears to be the returning of change during a business transaction. Event (13) aids in constructing a new model, namely, going dutch on a date to the movies. The Collins et al. (in press) examples illustrate what is meant by an interactive model (see Rumelhart, 1977 for a discussion of these kinds of models). The central point, though, in the examples is that inductive reasoning is initially involved in constructing the model. Once constructed, the process becomes top-down.

Once a model is constructed it enables the prediction of a number of events, including probable pre-conditions, causes and consequences of actions, emotional reactions, goals, etc. Those predictions are what guide the assimilation of new information into old and underlie the intense current interest in schemata (Bartlett, 1932), frames (Minsky, 1975), story grammars (Mandler & Johnson, 1977; Rumelhart, 1975; Stein & Glenn, 1979), scripts (Schank & Abelson, 1977) and other organized knowledge bases for comprehension.

In the "Mary" examples, when one combines events (7) and (9), a precondition is that the meal was prepared; an inference is that Mary was hungry and likes lamb; and a prediction is that since her dress is soiled her mother may become angry and Mary might be punished.

What Is Required to Make Inferences?

It is clear that background knowledge is needed to make inferences. What the reader knows or has experienced prior to reading a text is critical, and the reader's knowledge of the world or procedural knowledge may be decomposed into a number of knowledge domains. One implication of this is that if we want to enhance a child's comprehension of what he reads, we would do well to increase his general knowledge and understanding as well as teach him specific reading skills, after he has learned to decode. In addition, vocabulary (conceptualization) knowledge, regardless of domain, is a crucial pre-condition to comprehension (Pearson & Johnson, 1978; Trabasso, in press) since without understanding the basic concepts contained in the text or question, one cannot make inferential links.

Knowledge of text structure also helps comprehension. In narrative and expository text, this may aid in a top-down fashion. For example, since stories have well defined episodic structures (Mandler & Johnson, 1977; Stein & Glenn, 1979) the reader may establish expectations of settings, of events which create goals for the protagonist, of plans for achieving goals, of actions, of consequences or goal realizations and of reactions by the protagonist. These structures also presuppose context and relational, as well as functional, knowledge of the grammatical categories.

Knowledge about social interaction and human intentionality may aid comprehension. Stories entail considerable knowledge about social and personal interaction (Schantz, 1975) as well as about goals, plans, and actions (Schank & Abelson, 1977). In short, they represent a kind of naive psychology based upon a theory of actions and motives behind actions. Children acquire and use these naive theories of human motivation and goals to understand narratives. The problem is to determine what they know at different levels of development and how this knowledge interacts with what they read.

Finally, knowledge of causal relations between events is crucial for making inferences. The reader's ability to generate causes and consequences of events enables the prediction and assimilation of events into a causal chain representation as well as the filling in via inferences of missing information. With repeated exposure to situations, the reader develops stereotyped generalized experiences, called scripts (Schank & Abelson, 1977), which allow a well-constructed, known causal chain to predict behavior. Deviations from the script require further inferencing. When scripts are not available, the reader uses "plans" to acquire information and construct new scripts. In short, the reader's knowledge base, including his cultural background, appears to be the bottom line for comprehension.

What Processes Are Involved?

In the above discussion, reference was made to "top-down" and "bottom-up" processes interacting in making inferences. When "top-down," the construction of a causal chain, inferential prediction, and event integration is preceded by organized knowledge structures. When these are absent, the reader must use word recognition, word knowledge, and linguistic skills to derive sentence meanings and infer a model or framework.

How these processes are accomplished is a mystery, although some computer models are available, such as those discussed in Schank and Abelson (1977) or Kintsch and Van Dijk (1978). These approaches both involve the linking of propositions: In the Schank and Abelson's system the concepts are underlying meanings of arguments in propositions and the links are causal in nature whereas, in the Kintsch and Van Dijk approach, the linkages are determined by concept overlap or repetition across propositions.

What Kinds of Inferences Are There?

In this section, we shall briefly summarize the kinds of inferences detailed in an inference taxonomy by Warren, Nicholas, and Trabasso (1979). According to Warren et al. (1979), inferences may be divided into those which are logical (causal), those which are informational, those which are spatial or temporal, those which are related to script knowledge, those which depend upon world knowledge in some general sense, and those which are primarily evaluative in nature. Here, we shall stress the first class of inferences since we regard them as necessary to the construction of relations between events and the building of a causal chain representation. Informational inferences are thought to be more intrapropositional in nature, and while they are crucial to forming conceptualizations of sentence content, and precede the causal connecting of such conceptualizations, the construction of a causal chain is more important to what we are considering.

Logical inferences can go either in a forward (consequent) or backward (antecedent) manner. For example, if we know the goal of a protagonist, we can expect or predict certain actions to occur as consequences. On the other hand, knowing his actions constrains inferences about the reasons why he is doing what he does. Warren et al. (1979) distinguished among four types of logical relations.

1. *Motivational.* Goals *motivate* either other goals or such overt actions as events (goals also motivate cognitions [thoughts] and emotional reactions motivate goals and cognitions).

2. *Psychological causes.* Actions which are involuntary, as well as thoughts and feelings, are psychologically caused. Crying, inferring, and becoming angry are examples.

3. *Physical causes.* Physical or natural events or physical actions cause (mechanically cause) changes in state. Breaking a leg or drinking a glass of water are examples of actions which physically cause a change in state.

4. *Enablement.* Enablements are those conditions, typically states, which are necessary but not sufficient for a state or an action to occur. Having money enables one to buy things.

This listing of causal links resembles, in part, that of Schank and Abelson (1977). In their system, actions *result* in (physically cause) states, states *enable* acts, states or actions *initiate* (psychologically cause) a mental state, and mental actions (goals, thoughts, cognitions) are the *reasons for* (motivate) physical actions. In addition, one can have preventative causes where a state *disables* an action.

The logical relations identified above determine the kind of inferences made. If one focuses on an event and asks a why question about that event, then the kind of inference required is determined by the nature of the link. This does not mean, however, that the kind of processes invoked differ. The same process of finding events related to other events may occur for all four types. In fact, Omanson, Warren, and Trabasso (1978), using probe tests on children 5 to 8 years in age, failed to find consistent differences among logical causes.

How Can We Assess What Inferences a Reader Makes?

In this section, we shall not deal with the question of what inferences readers make during reading. At the moment, there are no adequate methods for assessing this (see Trabasso & Nicholas, 1980, for a review on inferences by children) and there is considerable debate about how many inferences are necessary for the construction of a representation of events in a narrative (Warren, Nicholas, & Trabasso, 1979). Therefore, a consideration of the kinds of question that could help in finding out if the reader *could* make certain inferences is now what needs to be discussed. In this discussion, we shall rely heavily upon a recent book by Lehnert (1978) on answering questions.

In order to illustrate the question types and relate them to the inference types above, read the Farmer and the Donkey story in Table 1.

If the reader generates inferences which result in the construction of a causal chain of events, then his ability to answer questions about logical relations either during or after the reading of a narrative should reflect this generative capability. Questions can be posed which assess the reader's knowledge of causal antecedents or causal consequents (Lehnert, 1978). The question itself contains a conceptualization, and the syntactic form of the question determines which kind of relation is being queried.

Referring to the Farmer and the Donkey story in Table 1, we can ask *causal antecedent* questions on inferences of the types previously described. Consider, first, the following variants of a physical causal antecedent question.

14. Why did the dog begin to bark loudly?
15. What caused the dog to begin to bark loudly?
16. What happened that resulted in the dog's beginning to bark loudly?
17. The dog barked loudly because...?

Table 1
The Farmer and the Donkey Story

1. There was once an old farmer
2. who owned a very stubborn donkey.
3. One evening the farmer was trying to put his donkey into its shed.
4. First, the farmer pulled the donkey,
5. but the donkey wouldn't move.
6. Then the farmer pushed the donkey,
7. but still the donkey wouldn't move.
8. Finally, the farmer asked his dog
9. to bark loudly at the donkey
10. and thereby frighten him into the shed.
11. But the dog refused.
12. So then, the farmer asked his cat
13. to scratch the dog
14. so the dog would bark loudly
15. and thereby frighten the donkey into the shed.
16. But the cat replied,
17. "I would gladly scratch the dog
18. if only you would get me some milk."
19. So the farmer went to his cow
20. and asked for some milk
21. to give to the cat.
22. But the cow replied,
23. "I would gladly give you some milk
24. if only you would give me some hay."
25. Thus, the farmer went to the haystack
26. and got some hay.
27. As soon as he gave the hay to the cow,
28. the cow gave the farmer some milk.
29. Then the farmer went to the cat
30. and gave the milk to the cat.
31. As soon as the cat got the milk,
32. it began to scratch the dog.
33. As soon as the cat scratched the dog,
34. the dog began to bark loudly.
35. The barking so frightened the donkey
36. that it jumped immediately into its shed.

Although why questions signal a causal antecedent relation, examples (14-17) indicate what question as well as what verbs or connectives can mark their relations. Note also that in each example, the same conceptualization (the dog barking loudly) is indicated. The reader then must search his memory for that conceptualization (here, the cat scratched the dog) *which resulted* in the dog being in pain (an inference) and his barking.

An example of a psychological antecedent causal question is given in (18).

18. Why did the barking frighten the donkey?

If an event leads to another event, and we pose questions about the first event, then we are asking for answers which call for *consequential conceptualizations.*

19. What happened when the farmer gave the cat milk?

20. What resulted from the farmer giving the cat milk?

21. What happened after the farmer gave the cat milk?

Examples (19-21) show variations on a causal consequence question concerning the goal satisfaction of the cat as a precondition for the cat scratching the dog. In general, consequence questions are signalled by what happens when . . .?

It is also possible to pose consequence questions negatively to see if the reader understands events that would not have occurred if certain pre-conditions weren't met or if certain antecedent events had not occurred. In the context of a story, these are hypothetical non-events. For example,

22. What if the farmer hadn't given the cat milk?

23. What would have happened if the farmer hadn't given the cat milk?

24. If the farmer hadn't given the cat milk, then what would have happened?

We can direct the reader towards consequences by providing information as in

25. What did the cat do after the farmer gave the cat milk?

Question (25) specifically directs the reader to the cat's action.

Motivational questions (what Lehnert, 1978, refers to as Goal Orientation) may be posed as antecedent or consequent (purpose) questions. For example,

26. Why did the farmer ask the cat to scratch the dog?

could be anwered by an antecedent event,

27. The dog refused to bark at the donkey.

or by a purpose,

28. He wanted to get the dog to bark in order to frighten the donkey and make him jump into the barn.

The event described in (27) resulted in or motivated the farmer to ask the cat to scratch the dog since it was a failure in his initial attempt at his superordinate goal of getting the donkey into the barn. That led to the farmer's subsequent actions. The event in (28) is the reason for or purpose of the farmer's asking the cat to scratch the dog.

Motivation questions may require answers involving more than one subgoal and a major goal. Examples (29), (30), and (31) contain questions on actions which could be answered by two, three, or four goals or motives, respectively.

29. Why did the farmer ask the cat to scratch the dog? (two reasons)

30. Why did the farmer ask the cow for milk? (three reasons)

31. Why did the farmer give hay to the cow? (four reasons)

The fourth logical relation, that of *enablement*, is usually marked by How or What and calls for answers involving states or action which satisfy specific pre-conditions necessary for the event in the question to occur. Examples (32-34) show some variations:

32. How was the farmer able to get the cow hay?

33. What did the farmer need to do in order to get the cow hay?

34. What did the farmer do in order to get the cow some hay?

Enablements may involve a long string of acts. If so, these become instrumental or procedural questions according to Lehnert (1978). For example, asking someone for directions to a house or how to cook *coq-au-vin* requires a listing of actions and instruments. In the Farmer and the Donkey story, this amounts to almost retelling the story in response to the question:

35. What did the farmer do in order to get the donkey into the barn?

Questions on logical relations between events either assess or prompt the reader's generation of text-connecting or slot-filling inferences. The questions considered next also assess or promote inferential comprehension but they do so *within* sentences.

The first set of within proposition questions contains what Lehnert (1978) classified as *concept completion* questions. These questions require that the reader search his memory or the text for a missing component. These questions basically interrogate case relations (agents, instruments, etc.), of which some examples are:

36. Who gave the farmer some milk?
37. What frightened the donkey?
38. What did the cat reply when the farmer asked him to scratch the dog?
39. What did the donkey refuse to do?
40. Where did the farmer go to get the hay?
41. When did the cat scratch the dog?
42. To whom did the farmer give the milk?

Questions (36-42) are probe questions. These could also be written as forced-choice, *disjunctive* questions such as,

43. Did the farmer give the milk to the dog or to the cat?

or as *verification* questions requiring a yes/no answer as in,

44. Did the farmer give the cow milk?

Disjunctive and verification questions are easier since they specify the conceptualization fully and require only a direct match between what is in memory and what is in the question. Furthermore, since they do not require a search among a large set of alternatives, the guessing probability is limited to one over the number of alternatives specified in the disjunction or one-half in the case of two alternatives and verification questions. The disjunctive, verification, and what Lehnert (1978) calls "feature specification" (e.g., What color is the dog?) questions are close to what is normally termed "literal comprehension." However, this term is misleading since even the understanding of sentences and their translation into a conceptualization involve considerable linguistic, semantic, contextual, and intential knowledge. It would be better to call these questions text-constrained and within propositional, given the theoretical framework used here.

Two other kinds of questions which cover several events that are inferential in nature also deserve mention. Both are judgmental in that they involve internal scales, one using social or personal opinion criteria, the other using quantification. For example,

45. What should the farmer have done to persuade the donkey to get into the barn?

calls for an opinion and for the generation of an alternative goal plan. The question involves an evaluation of what the protagonist did. While morality does not enter into this example, moral judgment questions are similar in form to (45).

The second type of question calls for quantification and entails knowledge of classes and class-inclusion relations or an underlying scale for a state. For example,

46. How many animals were there in the story?

47. How badly did the farmer want the donkey to get into the barn?

48. How did the donkey feel?

Can We Promote Comprehension through Asking Inferential Questions?

There has been a long history of study on whether asking adjunct questions before, during, or after reading helps reading comprehension (Anderson & Biddle, 1975). The answer seems to be that such questions may help or hinder and it is not clear as to why. Another question arises as well, namely whether we *promote or assess comprehension via questions*? One problem with prior research on this question is that the questions used were generated largely on intuitive and informal grounds and did not follow from a model for language comprehension. In this section, we shall explore some implications of the causal-chain theory.

According to the causal-chain model, the reader understands a narrative by 1) forming conceptualizations of sentences and 2) linking conceptualizations by generating inferences which connect them. Once the causal-chain is represented in memory, the reader is said to have understood the narrative and can now perform additional operations upon this representation by use of various interpretive or summarization or story grammar rules.

The formation of the underlying conceptualizations appears to be a necessary pre-condition to connecting them. Thus, developmentally, one might expect individual sentence comprehension to precede that of linking sentences via inferences. This, in fact, appears to be the case. Omanson, Warren, and Trabasso (1978) assessed within proposition comprehension of stories by five- and eight-year-old children by the use of concept completion questions. Then, they asked the same children to make logical inferences via the use of causal antecedent and motivational questions, the inferences involving the linkage of the same propositions which they had probed with concept completion questions. The five- and eight-year old children were matched on how well they answered the concept completion questions and then were compared on how well they answered the inference questions. The data showed two things: 1) as the children more accurately retrieved concept completion information, the percentage of correct inferences also increased—a result in line with the assertion that conceptual understanding underlies inference generation—but 2) the older children generated more correct inferences despite the fact the two age groups were matched with regard to their memory of the propositions upon which the inference was based. Thus, finding relations between conceptualizations increases with age, independent of the ability to form the conceptualization.

Returning to the question of comprehension assessment or promotion, we can now examine the possible influence of within and between conceptualization questions. In particular, if the reader is asked concept completion questions (who? whom? what?) after each action in the Farmer and the Donkey story, we can assess how well the reader understands individual propositions. It is possible that such questioning could promote sentence comprehension but not promote linking conceptualizations across sentences. In contrast, we could ask inferential questions (why?) which assess the reader's comprehension of relations between propositions. It is possible that questions which require the finding of logical relations between events during reading could promote comprehension and memory by establishing more links in the causal chain.

Wimmer (1979) has performed a provocative study on these questions, using the Farmer and the Donkey story. Wimmer studied how well four- and eight-year-old children could answer questions while listening to the story and also how well they could later retell the story. He asked different groups of children why questions and who/whom questions after each action in the story. (Unfortunately, no control group was run where no questions were asked so we can't assess the effect of questions per se.)

Apparently, comprehension, as assessed by immediate recall of the story, was not affected by the kind of question asked since the respective percentages of propositions recalled by the why and who/whom groups were 38 and 39. On this measure, the kind of question asked did not aid comprehension, i.e., the construction of a better memory representation. (Perhaps delayed recall would have been more sensitive to the quality of the representation.)

However, the why questions seem to have assessed the children's ability to construct a causal-chain representation better than the who/whom questions. First, the correlation between accuracy on the why questions and recall of the story was significant and higher than that for the who/whom questions. The respective correlations were .77 ($p < .01$) and .40 ($p > .05$). However, since the level of performance on probe questions for the who/whom group (86 percent) was higher than that on the why questions (63 percent), the differences between the correlations could have been a result of restriction of range rather than question effects.

Another analysis, however, suggests that the why questions assessed individual differences in comprehension better than the who/whom questions and supports the assumption that understanding the concepts within a sentence precedes understanding of relations between sentences. Wimmer compared those four-year-old children who answered *all* questions correctly on their ability to recall the story. While the number of subjects was small, those children (n = 4) who answered all the why questions recalled 80 percent of the story propositions and those (n = 8) who answered all of the who/whom questions recalled 46 percent. Further, age differences in recall were nearly eliminated when the four- and eight-year-old children were matched on answering why questions, here the respective percentages (and numbers) were 80 percent (n = 4) and 93 percent (n = 17).

Thus, we have some indication that children understand individual sentences before they connect them inferentially and that understanding of the logical relations between sentences leads to better retention of a narrative. The question as to whether questions promote comprehension and which questions one should use remains unanswered by the two studies discussed here.

One goal of the above presentation has been to provide a framework in which to assess reading or listening comprehension via questions. The types of questions asked are systematically related to the types of relations that exist between states or actions in a narrative. The advantage of the present approach is that it indicates the kind of processing required by the reader in understanding concepts and relations between concepts in stories. Since teachers try nearly exclusively to use questions as their main means to assess comprehension (Durkin, 1977), a framework for systematic question asking which either promotes or assesses comprehension should prove to be a useful aid. Basic research on the value of systematic and theory-based questioning should also evaluate the usefulness of such procedures.

References

ANDERSON, R.C., & BIDDLE, W.B. On asking people questions about what they are reading. In G.H. Bower (Ed.), *Psychology of learning and motivation* (Vol. 9). New York: Academic Press, 1975.

BARTLETT, F.C. *Remembering: A study in experimental and social psychology.* Cambridge, Eng.: Cambridge University Press, 1932.

BRANSFORD, J.D., & McCARRELL, N.S. A sketch of a cognitive approach to comprehension. In W.B. Wimmer & D.S. Palermo (Eds.), *Cognition and the symbolic processes.* Hillsdale, N.J.: Erlbaum, 1975.

COLLINS, A., BROWN, J.S., & LARKIN, K. Inference in text understanding. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, N.J.: Erlbaum, in press.

DEWEY, J. *How we think.* Portions published in R.M. Hutchins & M.J. Adler (Eds.), *Gateway to the great books* (Vol. 10). Chicago:

Encyclopedia Britannica, 1963. (Originally published by Heath, 1933)

DURKIN, D. *Comprehension instruction—Where are you?* (Reading Education Report No. 1.) Urbana: University of Illinois, Center for the Study of Reading, October 1977. (ERIC Document Reproduction Service No. ED 146 566)

KINTSCH, W., & VAN DIJK, T.A. Toward a model of text comprehension and production. *Psychological Review*, 1978, *85*, 363-394.

LEHNERT, W.G. *The process of question answering.* Hillsdale, N.J.: Erlbaum, 1978.

MANDLER, J.M., & JOHNSON, N.S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 111-151.

MINSKY, M. A framework for representing knowledge. In P.H. Winston (Ed.), *The psychology of computer vision.* New York: McGraw-Hill, 1975.

NICHOLAS, D.W., & TRABASSO, T. Towards a taxonomy of inferences. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children.* Hillsdale, N.J.: Erlbaum, in press.

OMANSON, R.C., WARREN, W.H., & TRABASSO, T. Goals, inferential comprehension, and recall of stories by children. *Discourse Processes*, 1978, *1* (4), 337-354.

PEARSON, P.D., & JOHNSON, D.P. *Teaching reading comprehension.* New York: Holt, Rinehart & Winston, 1978.

RUMELHART, D.E. Notes on a schema for stories. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

RUMELHART, D.E. Toward an interactive model of reading. In S. Dornic (Ed.), *Attention and performance VI.* Hillsdale, N.J.: Erlbaum, 1977.

SCHANK, R.C. The structure of episodes in memory. In D.G. Bobrow & A.M. Collins (Eds.), *Representation and understanding: Studies in cognitive science.* New York: Academic Press, 1975.

SCHANK, R.C., & ABELSON, R.P. *Scripts, plans, goals and understanding.* Hillsdale, N.J.: Erlbaum, 1977.

SHANTZ, C.U. The development of social cognition. In E.M. Hetherington (Ed.), *Review of child development research* (Vol. 5). Chicago: University of Chicago Press, 1975.

STEIN, N.L., & GLENN, C.G. An analysis of story comprehension in elementary school children. In R.O. Freedle (Ed.), *New directions in discourse processing.* Hillsdale, N.J.: Erlbaum, 1979.

TRABASSO, T. Can we integrate research and instruction on reading comprehension? In C. Santa & B. Hayes (Eds.), *Children's prose comprehension: Research and practice.* Newark, Del.: International Reading Association, in press.

TRABASSO, T., & NICHOLAS, D.W. Memory and inferences in the comprehension of narratives. In F. Wilkening, J. Becker, & T. Trabasso (Eds.), *Information integration by children.* Hillsdale, N.J.: Erlbaum, in press.

WARREN, W.H., NICHOLAS, D.W., & TRABASSO, T. Event chains and inferences in understanding narratives. In R. O. Freedle (Ed.), *New directions in discourse processing (Vol. 2).* Hillsdale, N.J.: Erlbaum, 1979.

WIMMER, H. *Children's comprehension and recall of hierarchically structured stories.* Paper presented at the meetings of the Society for Research in Child Development, San Francisco, March 1979.

Vocabulary Knowledge

Richard C. Anderson
Peter Freebody
University of Illinois at Urbana-Champaign

Our aim in this paper is to summarize what is known about the role of vocabulary knowledge in reading comprehension. Though word identification skills are important in reading, this paper is concerned exclusively with *knowledge of word meanings*. An assessment of the number of meanings a reader knows enables a remarkably accurate prediction of this individual's ability to comprehend discourse. Why this is true is poorly understood. Determining why is important because what should be done to build vocabulary knowledge depends on why it relates so strongly to reading. The deeper reasons why word knowledge correlates with comprehension cannot be determined satisfactorily without improved methods of estimating the size of people's vocabularies. Improved assessment methods hinge, in turn, on thoughtful answers to such questions as what is a word, what does it mean to know the meaning of a word, and what is the most efficient way of estimating vocabulary size from an individual's performance on a sample of words.

Vocabulary Knowledge and Linguistic Ability

Measures of vocabulary knowledge are potent predictors of a variety of indices of linguistic ability. The strong relationship between vocabulary and general intelligence is one of the most robust findings in the history of intelligence testing. Terman (1918), for instance, reported a correlation of .91 between mental age (as assessed by the Stanford Revision of the Binet-Simon Scale) and the vocabulary subscale. On this basis he suggested that the vocabulary measure alone constitutes a good estimate of performance on the entire scale and thus could be used as a short measure. Since then, this suggestion has been tested with various age groups. Table 1 summarizes representative evidence. In these studies, correlations between vocabulary subtest scores and total test scores on a number of different IQ and achievement tests have ranged from .71 to .98.

An equally consistent finding has been that word knowledge is strongly related to reading comprehension. Davis (1944a, 1968) factor analyzed nine comprehension tests and found a main factor for word knowledge on which a vocabulary test loaded about .8. Thurstone (1946) reanalyzed Davis' original data and found three major factors: vocabulary knowledge, ability to draw inferences from a paragraph, and ability to grasp the main idea of a paragraph. In the years that followed, several factor analytic studies identified a "reading comprehension" factor (Botzum, 1951; Clark, 1972; Fruchter, 1948; Wrigley, Saunders & Newhaus, 1958). The range of factor loadings for vocabulary tests in these studies was .41 to .93.

This strong relationship has been found to hold across a wide range of language groups. Thorndike (1973) collected data from over 100,000 students from 15 countries, across three age groups; he found median correlations between vocabulary knowledge and reading comprehension, corrected for test reliability, of .71 (10-year-olds), .75 (14-year-olds), and .66 (17-18-year-olds). Thorndike concluded that the results indicate "how completely reading performance is determined by word knowledge at different levels and in different countries" (p. 62). The uncorrected correlations are reproduced in Table 2.

Analyses of readability (cf. Bormuth, 1966) also demonstrate the preeminent role of word knowledge. In a study of the factors that make prose difficult to read, Coleman (1971) examined morphological, syntactic, and semantic properties of words and sentences. While he found sentence complexity to be a fairly important variable, he was able to conclude that "any measure of word complexity (number of letters, morphemes, or

Table 1
Correlations of Various Vocabulary Tests with Tests of General Intelligence

Vocabulary Measure	Intelligence Measure	Subjects	*N*	*r*	Source
Terman, 1916	Binet (1916)	School children	631	.91	Terman (1918)
Terman, 1916	Binet (1916)	School children	269	.87	Mahan & Witmer (1936)
Terman, 1937	Binet (1916)	School children	65	.92	Spache (1943)
Terman, 1937	Binet (1916)	School children	1161	.98	Elwood (1939)
Terman, 1937	Binet (1916)	School children	753	.86	White (1942)
Terman, 1937	Binet (1916)	Standardization sample, ages 8-18	710	.71 to .86	McNemar (1942)
Wechsler	Wechsler	Adult males	1000	.82	Lewinski (1948)
Wechsler	WISC	Standardization sample, ages 7.5, 10.5, 13.5	600	.71 .87 .78	Wechsler (1949)
Raven	Binet	School children	150	.93	Raven (1948)
Dupuy	Various tests	School children	2397	.76	Dupuy (1974)
	Stanford Achievement Tests (1973)	Standardization samples			Stanford Achievement Tests (1973)
		Grade 2		.82	
	(vocabulary with total achievement test scores)	3		.79	
		4	275,000 over grades and geog. locale	.80	
		5		.80	
		6		.83	
		8		.89	

Note. Adapted from Miner (1957).

Table 2
Correlations of Word Knowledge with Reading Comprehension
in Fifteen Countries

Country	Age Group		
	10 yrs.	14 yrs.	17 yrs.
Belgium (Fl)	.537	.591	.500
Belgium (Fr)	.588	.619	.481
Chile	.543	.508	.577
England	.735	.698	.497
Finland	.617	.654	.395
Hungary	.594	.533	.389
India	.569	.387	.320
Iran	.498	.427	.294
Israel	.651	.674	——
Italy	.580	.587	.446
Netherlands	.620	.624	.310
New Zealand	——	.685	.536
Scotland	.716	.719	.579
Sweden	.559	.598	.584
United States	.735	.693	.679

Note. From Thorndike (1973).

syllables; frequency of usage) will account for about 80 percent of the predicted variance" (p. 184). Klare (1974-1975), in a review of readability, also concluded that a two-variable formula is sufficient for most practical purposes: one variable relates to word difficulty and the other to syntactic or sentence difficulty. He went on to conclude that the word variable is consistently more highly predictive of difficulty than is the sentence variable. As would be expected, some index of vocabulary difficulty has typically been given the heaviest weight in readability formulas.

Why is Vocabulary Knowledge a Major Factor in Linguistic Ability?

There are three more or less distinct views of why vocabulary knowledge is such an extraordinary correlate of linguistic ability. We will call the first the *instrumentalist* position: individuals who score high on a vocabulary test are likely to know more of the words in most texts they encounter than low scoring individuals. The heart of the instrumentalist hypothesis is that knowing the words enables text comprehension. In other words, this hypothesis claims that vocabulary knowledge is directly and importantly in the causal chain resulting in text comprehension. Unlike the two positions described below, the instrumentalist hypothesis has nothing to say about where vocabulary knowledge comes from but only that, once possessed, it helps the reader understand text.

According to the second position, vocabulary tests measure verbal *aptitude*. A person who scores high on such a test has a quick mind. With the same amount of exposure to the culture, this individual has learned more word meanings. He or she also comprehends discourse more readily than the person who scores low on a vocabulary test. The essential claim of the aptitude hypothesis is that persons with large vocabularies are better at discourse comprehension because they possess superior mental agility. A large vocabulary is not conceived to be involved in a direct way in better text understanding in this model. Rather vocabulary test performance is merely another reflection of verbal ability and it is verbal ability that mainly determines whether text will be understood.

The third position is the *knowledge* hypothesis. Performance on vocabulary tests is seen as a reflection of the extent of exposure to the culture. The person who scores high has deeper and broader knowledge of the culture. The essential idea is that it is this knowledge that is crucial for text understanding. Rather than being directly important, possessing a certain word meaning is only a sign that the individual may possess the knowledge needed to understand a text. For instance, the child who knows the word *mast* is likely to have knowledge about sailing. This knowledge enables that child to understand a text that contains sentences which do not even involve the word *mast*, such as, "We jibed suddenly and the boom snapped across the cockpit."

Of course, *jibe, boom*, and *cockpit* are specialized words, too. It might be wondered whether the instrumental hypothesis and the knowledge hypothesis are really different. Strong versions of the two positions are distinguishable, at least. The instrumental position, as we choose to characterize it, stresses individual word meanings. The knowledge view emphasizes conceptual frameworks or "schemata"; individual word meanings are merely the exposed tip of the conceptual iceberg.

Which of these three positions is most tenable? The main point to be made is that there are neither the theoretical tools nor the data to justify a conclusion at the present time. A second important point is that it would be naive, indeed, to assume that one of the positions will turn out to be entirely right and the other two entirely wrong.

The most fully developed position is that vocabulary knowledge reflects verbal aptitude. As the studies reviewed earlier indicate, vocabulary tests intercorrelate highly with a variety of other kinds of tests reflecting "intelligence." On its face, this fact is hard to understand solely in terms of the instrumentalist or knowledge positions. Probably by metaphorical extension of notions of physical agility, it is customary to speak of people of high intelligence as having "quick" minds. Recently Earl Hunt and his associates have been trying to prove that this is more than a metaphor (cf. Hunt, 1978). They theorized that people of high verbal ability are literally faster than other people at elemental verbal coding and recoding operations. One task used to assess speed of mental operations developed by Posner (cf. Posner & Mitchell, 1967) involves the subjects' deciding whether pairs of upper or lower case letters match. In one condition, the subject has to judge if two letters have the same name (e.g., aA), and in the other condition, the decision is whether or not the letters are physically identical (e.g., AA). The subjects' responses are timed. It is argued that a time measure derived from this task is a pure index of the speed of some elemental verbal operations, since the subject needs to "look up" in memory the names of the two letters and compare them. Hunt and his collaborators have found that this measure correlates about .30 with standardized tests of verbal ability. This is a relationship that could not have been predicted and is not readily explained by the other hypotheses being entertained.

Nevertheless, the case is far from conclusive. The general ability tests used in Hunt's studies probably placed subjects under at least some implicit time pressure. This could have given fast workers an advantage. If so, the studies may have revealed that fast people are fast rather than that fast people are smart. Consistent with this interpretation are the results of a factor analysis of representative paper and pencil ability measures and laboratory reaction time tasks completed by Hunt, Lunneborg, and Lewis (1975). The measures of speed of really elemental processes, such as letter matching time, loaded on a factor that appears to represent clerical speed and accuracy instead of on the factor representing general intelligence. A study of Kirby and Das (1977) also indicated that processing speed is a separable factor in tests of verbal and spatial abilities. This conclusion seems to be a sound one since Thorndike (1973) found, in his study of reading comprehension in 15 countries, only modest correlations between performance on reading speed and comprehension tests. The median corrected correlations were .42 for 10-year-olds and .47 for 14-year-olds.

With respect to the instrumentalist position, as the evidence reviewed earlier indicates, word difficulty is highly predictive of readability. Does this fact clinch the argument in favor of the instrumental hypothesis? No, since it is possible that variation among texts in vocabulary difficulty is merely symptomatic of deeper differences in knowledge prerequisites. To prove that knowing the meaning of individual words has an important instrumental role in understanding text would require more than correlational evidence. It would need to be shown a) that the substitution of easier or more difficult words in a text makes that text easier or more difficult to comprehend, and b) that people are helped to comprehend a text if they learn the meanings of the unfamiliar words it contains. A cursory look at the literature bearing on these points suggests that the assumptions of the instrumentalist position are unquestioned tenets rather than hypotheses in need of verification.

There is some research in which texts have been altered so as to vary word familiarity (see Chall, 1958, for a review of the early studies). In a recent set of experiments, Wittrock, Marks, and Doctorow (1975; see also Marks, Doctorow, & Wittrock, 1974) replaced 15 percent of the words in several passages with either high-frequency or low-frequency synonyms.

There is some confusion in the Wittrock et al. paper over the word frequency manipulations. While this detracts from the findings, the conclusions may hold for "easy" and "hard" words. Sixth graders of every level of reading skill evidenced better comprehension of texts containing easy words than texts containing hard words, whether they were reading or listening. Furthermore, children who began with an easy text later showed improved comprehension of the hard version of the same text. Performance on a vocabulary test suggested that children who had first received the easy version of a passage were able to learn some of the low-frequency words in the hard version.

Other recent evidence is less favorable to the instrumentalist position. Tuinman and Brady (1974) were unable to increase fourth, fifth, and sixth grade students' comprehension of texts that contained a substantial proportion of difficult words by direct instruction on those words, even though such instruction significantly increased the students' performance on the vocabulary items themselves. These authors concluded that the instrumental hypothesis seems to be ruled out. Jenkins, Pany, and Schreck (1978; see also Pany & Jenkins, 1977) were also unable to establish that vocabulary instruction improves reading comprehension. Several different methods for teaching word meanings were explored. All were at least somewhat better than no instruction. The method which proved most effective with both average and learning disabled children involved intensive drill and practice on the words in isolation. However, even when children had definitely learned the meanings of twelve difficult words, they did no better than uninstructed children who definitely did not know these words on a cloze test or in retelling a brief story containing the twelve difficult words. We do not know how to reconcile the conflicting results bearing on the instrumental hypothesis other than to conclude, as reviewers of educational research must so often conclude, that more research is needed.

Turning now to the third position, there is now a truly substantial case that background knowledge is crucial for reading comprehension (cf. Anderson, 1978). However, there is thin evidence to support the view that vocabulary scores primarily reflect such background knowledge. We shall cite just one study which suggests that the idea is plausible. Steffensen, Joag-Dev, and Anderson (1979) asked natives of the U.S. and of India to read passages describing an American and an Indian wedding. The results showed that the native passages were read more rapidly and recalled in greater detail. There were more culturally appropriate elaborations of the native passages and more culturally inappropriate distortions of the foreign ones. The vocabulary of the two passages was closely controlled. For instance, there were only two words in the Indian passage, *sari* and *dhoti*, referring to articles of women's and men's clothing, respectively, that would have been unfamiliar to any of the American subjects. These two words did not figure in any important way in the passage, so failure to know them could have had no more than a negligible effect. Still, a two item vocabulary test, examining knowledge of *sari* and *dhoti*, would have been an excellent predictor of performance on the Indian passage. All Indian subjects would have known both words. Some Americans would have known *sari* but very few would have known *dhoti*. It is apparent that the test would have neatly divided subjects in terms of the extent of their knowledge of Indian culture, which was obviously the underlying reason for the large observed differences between Indians and Americans in comprehension, learning, and memory.

Instructional Implications of Different Hypotheses about Vocabulary Knowledge

It is important to know which of the three hypotheses about vocabulary knowledge is most nearly correct because the views have radically different implications for the reading curriculum. At one extreme, some who endorse the verbal aptitude hypothesis are fatalistic about whether any environmental factor can have a major influence on children's reading. They tend to recommend family planning instead of curriculum innovation as the final solution to the reading problem. Of course the verbal aptitude position does not require the belief that heredity is predominant. Alternatively, there are those who maintain that verbal ability grows in proportion to the volume of experience with language. The greater the opportunities to use language the faster and more efficient become the elemental processing operations. In turn, speed and efficiency permit greater benefit from each successive language encounter. More detailed accounts of this sort of position can be found in the well-known paper by LaBerge and Samuels (1974) and a recent paper by Perfetti and Lesgold (1979).

The latter formulation of the verbal aptitude hypothesis leads to the recommendation that educators should try to maximize the amount of reading children do. However, this is not very newsworthy. It is a practice that would be endorsed no matter what the theoretical persuasion. The distinctive emphasis in the verbal aptitude position is on speed and efficiency of processing. This emphasis gives rise to the recommendation that beginning readers and poor readers receive extensive drill and practice on "fundamentals" of reading. According to Perfetti and Lesgold (1979), the drill activities should include even more practice than typically provided in word vocalization, more practice in speeded word recognition, and more practice in immediate memory for the literal content of text. It should be noted that these suggestions are offered in the spirit of a hypothesis. Perfetti and Lesgold acknowledge that, so far at least, attempts to facilitate text comprehension by providing speeded word drills have not proved very successful (see especially Fleisher & Jenkins, 1977).

While, like everyone else, the advocate of the instrumental hypothesis favors lots of reading and varied language experience, the distinctive feature of this view is that it invites direct vocabulary building exercises. Becker (1977) has argued strongly for the instrumentalist position. He maintained that once decoding skills have been mastered, the chief remaining factor in determining whether a child will be a successful reader is vocabulary knowledge. He claimed that schools have never had reading programs that systematically build vocabulary. Children from middle class backgrounds pick up word meanings anyway. But the same is less true, Becker argued, of children coming from lower class homes, which often fail to provide support for the continuous vocabulary and concept growth important to school work. Consistent with this assumption is some recent work by

Hall and Tirre (1979), who found that lower class parents, particularly lower class black parents, use substantially fewer of the words found in standardized intelligence tests when speaking with their children than do middle class parents.

Becker proposed a reading curriculum in which every child would learn about 7,000 basic words from direct instruction. The figure 7,000 comes from one estimate of the number of basic words known by the average high school senior (Dupuy, 1974). Becker acknowledged that there are families of words with related meanings, thereby permitting the child some generalization beyond the words that are specifically taught. By and large, though, he believed that learning one vocabulary item gives little advantage in learning the next one. For instance, he illustrated morphological instruction on the following set of unrelated words: *help, support, insist, toil, resist, recognize, assist*. Even his so-called "concept side" of the instruction entailed a component analysis of isolated words. So if this assumption is correct, direct teaching of a vocabulary of even 7,000 basic words would be an enormous task. Becker estimated that about 25 basic words would have to be taught per week from the third through the twelfth grade (p. 539).

The distinctive curriculum implication of the knowledge hypothesis is that generally new vocabulary ought to be learned in the context of acquiring new knowledge (cf. Goodman, 1976, p. 487). Every serious student of reading recognizes that the significant aspect of vocabulary development is in the learning of concepts not just words. The additional point that the knowledge position brings to the fore is that concepts come in clusters that are systematically interrelated. Returning to an earlier example, the concept of *mast* cannot be acquired independently of concepts such as *boat* and *sail*. Thus, it would seem to be sensible for people to learn the jargon in the context of learning about sailing and the anatomy of sailboats. According to the knowledge hypothesis, if a child were really naive, trying to teach a single sailing concept and word in isolation from the set of related concepts and words would be inefficient in the best case and completely fruitless in the worst case.

A thought experiment suggests the more general point about the role of knowledge in vocabulary learning. Suppose you wished to teach some French vocabulary to, let us say, two groups of English-speaking Canadian children, evenly matched on aptitude and achievement. One group is from a downtown urban area, the other is from a small fishing village. The body of words you wish to teach is concerned with fishing (*trawlers, rods, nets, casting, bait, currents*, etc.). Would you expect one group to learn the words more quickly and easily than the other? Why? We do not know of research that has dealt systematically with these questions. One somewhat relevant study was carried out by Allen and Garton (1968). They found that physics students were much better than art students in recognizing physics words. They concluded that, for art students, physics words are semantically indistinct and thus have to be recognized on a more piecemeal basis. Familiarity with an area of knowledge increased the familiarity of the physics words.

Knowledge can be sliced in various ways. Thus far in this section, we have considered sets of words related because they are used in talking about the same topic. Words may also be conceptualized in terms of families related to one another because they convey related sets of distinctions. Consider an example involving verbs of visual perception.[1] The basic verb is *see*. If you notice that *look* involves a deliberate act of seeing, it can then be appreciated that *glimpse* refers to a short act of seeing whereas *glance* refers to a short act of looking. *Stare*, on the other hand, refers to a prolonged act of looking. The variations in sense among these verbs can be understood in terms of just two semantic features, intention and duration. Further distinctions would be required to encompass other verbs of visual perception such as *notice* and *examine*.

We would consider that a lesson that helped children sharpen and extend the distinctions involved in visual perception words to be consistent with the spirit of the knowledge position. What the knowledge position would not countenance is a separate vocabulary lesson that included *glance, mast*, and a miscellany of other words. Herein lies a difference from the instrumentalist position, which does not seem to us to preclude exercises involving lists of unrelated words.

---

[1] We are indebted to Charles Fillmore for this example.

Johnson and Pearson's (1978) book, *Teaching Reading Vocabulary*, appears to represent predominantly the knowledge position, though it is an eclectic treatment that also reflects influences from the other two views. Johnson and Pearson advocated teaching a basic sight vocabulary using "intensive direct instruction in the early grades and with older children who do not read well!" (p. 28). They also endorsed both direct and indirect means for teaching phonics, promoting morphological analysis, causing vocabulary knowledge to expand, and teaching the use of the dictionary and thesaurus. Johnson and Pearson devoted a chapter to the use of contextual clues to figure out the meanings of unfamiliar and ambiguous words. Otherwise most of the exercises and games suggested throughout the book involve sets of words outside the context of stories or textbook chapters. However, the words usually involved sets of interrelated distinctions, such as were illustrated above with verbs of visual perception. Almost every activity was designed to expand children's sensitivity to these distinctions. There is an apparent discrepancy between the goals of the activities, which are concerned with conceptual distinctions and relations, and the format of the activities, which is based largely on isolated words. If the knowledge perspective were strictly adhered to, vocabulary instruction would not be thought of as a separate subject in school.

For the sake of clarity of exposition, we have presented the aptitude, instrumental, and knowledge positions in uncomplicated and somewhat overdrawn form. We must emphasize again that no serious scholar in reading or related fields rigidly adheres to any one of these positions. In particular Hunt, who has been identified with the aptitude hypothesis, has explicitly and emphatically stated that vocabulary size also is a reflection of an individual's accumulated knowledge of the world. Becker, whom we labeled an instrumentalist, heartily endorses some of the implications of both the aptitude and the knowledge views. Reading has been a fractious field. If a policy were followed of avoiding controversy where none genuinely exists, the quality of intellectual exchange and the sociopolitical climate might improve to the point where someone within the next decade could write a book entitled "Learning to Read: The Great Consensus."

What Does It Mean to Know the Meaning of a Word?

It is not clear that, if Ludwig Wittgenstein and Bertrand Russell were left alone in a room for three hours, they could decide that they really knew the meaning of *dog*. As Labov (1973, p. 341) said, "Words have often been called slippery customers, and many scholars have been distressed by their tendency to shift their meanings and slide out from under any simple definition."

An ordinary adult engaging in an ordinary conversation will be absolutely sure he knows the meanings of almost all of the words he hears. Notice that the restriction to ordinary use is an important aspect of this confidence. Consider the term *gold*, for example. The person who is sure he knows the meaning of this word in an ordinary use will quickly retreat when in the company of jewelers, mining engineers, geological survey assayists, or metallurgists.

What does a person know when he knows the meaning of a word in its ordinary, everyday, garden-variety sense? This issue is addressed in what we will refer to as the Standard Theory of semantics, according to which the meaning of a word can be analyzed into features (also called components, attributes, or properties), each of which represents one of the distinctions conveyed by the word. Necessary or essential features are usually distinguished from features that are merely characteristic. For instance, having a back could be said to be a necessary feature of *chair* since an object that is otherwise a chair except for the lack of a back is really a stool instead of a chair. On the other hand, the ability to fly is only a characteristic feature of *bird* since some birds (penguins) don't fly at all and others (chickens) do so very poorly.

To define a term, in the strong sense, is to list the features necessary to capture the essence of the thing (or event or quality) designated by the word. Saying this another way, a proper definition indicates the attributes a thing must have in order to be designated by a word; if any of these necessary properties were missing that word would not apply. Before we choose this as our criterion in the testing of children's word knowledge, however, we might wish to examine how well it applies to adults' normal use and understanding of words.

How able are people to define the words they are sure they know? "Not very" is the answer if one insists upon the strong sense of define. Consider *gold* again. Upon being asked to define *gold*, the ordinary citizen might say that gold a) is precious, b) is a metal, and c) has a particular yellowish (i.e., golden) hue. The problem is that none of these is a necessary feature. Not all gold is a golden color. If, say, the Chinese were to discover a mountain of gold, the substance would no longer be precious. Not even the attribute of being a metal can be considered to be an eternal, immutable property of gold for, unlikely though it is, there might be a scientific breakthrough in which it was discovered that gold is not a metal.

A unicorn is a beast with such and such defining characteristics. Of course there are no beasts with these properties; which is to say that unicorns do not exist. By the same logic, if being precious and being a metal are defining features of gold, it follows that if the Chinese were to discover a mountain of the substance or scientists were to determine that the substance is not a metal, one would be forced to conclude that gold did not exist. As Putnam (1975) has noted, this is a very odd conclusion, because there would still be this "stuff" lying around that people used to call gold. We have a right to be suspicious of a semantic theory that backs us into such a peculiar corner.

Another example will illuminate the point even more starkly. When it comes to fine points of meaning, ordinary folks turn to experts as the final arbiters—to jewelers and metallurgists for the exact meaning of *gold*, to the Supreme Court for the proper interpretation of words in the Constitution, and so on. For the sake of the argument, it may be supposed that the American Psychiatric Association is the final arbiter of the meaning of *homosexual*. For years, this august group defined homosexuality as a disease of sexual orientation. Recently, however, the association declared that homosexuality is not a disease. Anita Bryant may not have agreed with that conclusion, but at least she understood it. If the characterization of homosexuality as a disease had been taken seriously as a *defining* feature, upon reconsidering its position, the American Psychiatric Association would have had to assert, "There is no such thing as homosexuality." That conclusion would simply have left Ms. Bryant puzzled.

There are other serious problems with Standard Theory. Notably, the members of a class called by the same name frequently do not all share a single set of common properties. Wittgenstein (1953; see also Rosch, 1973; Rosch & Mervis, 1975) argued that things designated by the same word generally are related by "family resemblance." He intended an analogy to a human family whose members look and act alike. Mother and one son may have a prominent nose. Father and daughter may have the same hair color. And so on. But there may be no single respect in which they are all alike, no single feature which they all share. Wittgenstein claimed family resemblance was the most accurate characterization of the relationships among the various uses of most common words. To illustrate his point, he analyzed uses of the term *game*, noting the similarities and differences between team games, board games, and children's games. Others have shown the fuzziness and context sensitivity of the meanings of terms such as *cup* (Labov, 1973); *eat* (Anderson & Ortony, 1975); *red* (Halff, Ortony, & Anderson, 1976); and *held* (Anderson, Pichert, Goetz, Schallert, Stevens, & Trollip, 1976).

A great deal more could be said about semantic theory. (For authoritative, current treatments, see Clark & Clark, 1977, especially chapters 11-14; Fillmore, 1975; and Miller & Johnson-Laird, 1976.) The main point of this brief excursion into the meaning of meaning is to caution against holding up a standard of word comprehension for children that adults could not meet.

Depth of Word Knowledge

It is useful to distinguish between two aspects of an individual's vocabulary knowledge. The first may be called "breadth" of knowledge, by which we mean the number of words for which the person knows at least some of the significant aspects of meaning. Later sections of this paper will be concerned mainly with breadth of knowledge.

Treated in this section is a second dimension of vocabulary knowledge, namely the quality or "depth" of understanding. We shall assume that, for most purposes, a person has a sufficiently deep understanding of a word if it conveys to him or her all of the distinctions that would be understood by an ordinary adult under normal circumstances.

Eve Clark (1973) has marshalled an array of evidence which shows that the meaning a young child has for a word is likely to be more global, less differentiated than that of an older person. With increasing age, the child makes more and more of the adult distinctions. In other words, when first acquired, the concept a child has for a word need not include all of the features of the adult concept. Eventually, in the normal course of affairs, the missing features will be learned.

While there are some differences in theoretical interpretation and some findings appear to hinge on procedural details (Brewer & Stone, 1975; Glucksberg, Hay, & Danks, 1976; Nelson, 1977; Richards, 1976), most of the research done to date supports the conclusion that there is progressive differentiation of word meanings with increasing age and experience.

Just one illustration will be provided of the kind of evidence that points to this conclusion. Gentner (1975) completed a theoretical analysis of verbs of possession which indicated that *buy*, *sell*, and *spend* entail a more complex set of distinctions than *give* and *take*. Notice that giving involves the transfer of something from one person to another. Selling likewise involves the transfer of something from one person to another, but it involves an additional transaction as well, the transfer of money from the buyer to the seller. The complementary relationship holds between buying and taking.

Gentner expected children to acquire the full, adult meanings of these verbs in order of complexity. Children ranging from four to eight years of age were asked to make dolls act out transactions from directions involving each verb. For example, the children were requested to "make Ernie sell Bert a (toy) car." The four-year-olds performed flawlessly with directions containing *give* and *take*, but never correctly executed instructions that involved *spend, buy,* or *sell.* The eight-year-olds exhibited nearly perfect understanding of every direction except the ones containing *sell.* Overall, the results were exactly as expected: the adult meanings of verbs of possession are acquired in order of complexity.

Gentner's analysis (1975, p. 242) of the children's errors suggests that the younger ones treated the complex verbs as though they were simpler forms. She explained, "... the commonest incorrect response was some form of one-way transfer... the young child acting out *buy* and *sell* completely disregards the money transfer that should be part of their meanings, yet performs the object transfer in the correct direction. He reacts to *buy* as if it were *take.* He treats *sell* as if it were *give.*" When asked to "make Bert spend some money," even the youngest child correctly handles the money transfer but neglects to have Bert get anything for the money he "spends." The child treats *spend money* as though it meant give money away.

Through some quirk of the sociology of science, the in-depth study of word knowledge has been the special province of psycholinguists studying language development in young children. There is a substantial body of literature on selected vocabulary of children from about two through eight years of age. The literature involving older children and adults is meager.

In our judgment, vocabulary knowledge continues to deepen throughout lifetimes; that is, as they grow older, most people continue to learn nuances and subtle distinctions conveyed by words that in some sense they have known since childhood. There is no hard data to support this conjecture. However, an illustration will show that many adults still have something to learn about even fairly common words. It is easy to find educated adults who confuse *infer* and *imply.* A person will say something along the lines, "I intended, by stating these arguments, to infer that...." Of course, this individual should have said *imply.* Speakers imply. Listeners infer. The complication, which no doubt makes the distinction difficult, is that speakers may report inferences they have made as well as get implications across to listeners.

Breadth of Word Knowledge

It is disturbing to examine available estimates of the average vocabulary size of various age groups. Table 3 summarizes studies that have been carried out to estimate total basic or "root" word knowledge. It can be seen that the estimates vary wildly.

It is not obvious how to evaluate the different sampling methods and response criteria that have been employed in research attempting to estimate vocabulary size. Recently, for instance, the distinguished psycholinguist, George Miller (1978), stated:

> Although the rapid rate of syntactic acquisition has inspired much respectful discussion in recent years, the rate of lexical growth is no less impressive. The best figures available indicate that children of average intelligence learn new words at a rate of more than 20 per day. It seems necessary to assume therefore, that at any particular time they have hundreds of words roughly categorized as to semantic or topical relevance but not yet worked out as to precise meaning or use. (p. 1003)

Miller did not specify whether or not he was referring to "basic" words. If he was, then he is positing a mean annual word acquisition rate of over seven thousand words, or about fifty thousand over the elementary and middle school years. This seems unlikely even in the light of the highest estimates summarized in Table 3. Miller may have been including compounds and derivatives. However, to our knowledge, no systematic examination of children's ability to understand these forms has been completed. Miller's statement highlights two points. First, in its original context, the statement is a crucial step in an argument about lexical development. Accurate estimates of the growth of word knowledge are an important element in discussions of lexical and conceptual development and the relationship between them. Second, how do we assess what are the "best figures available"?

In 1940, Seashore and Eckerson remarked that, even though the field of vocabulary testing is a "fairly old one" (p. 35),

Table 3
Some Previous Estimates of Total Vocabulary Size at Selected Grades

Grade	Source	Estimate
1st	M.E. Smith (1926)	2,562
	Dolch (1936)	2,703
	Ames (1964)	12,400
	M.K. Smith (1941)	17,000
	Shibles (1959)	26,000
3rd	Dupuy (1974)	2,000
	Holley (1919)	3,144
	Terman (1916)	3,600
	Brandenburg (1918)	5,429
	Kirkpatrick (1907)	6,620
	Cuff (1930)	7,425
	M.K. Smith (1941)	25,000
7th	Dupuy (1974)	4,760
	Terman (1916)	7,200
	Holley (1919)	8,478
	Kirkpatrick (1907)	10,666
	Brandenburg (1918)	11,445
	Cuff (1930)	14,910
	Bonser, et al. (1915)	26,520
	M.K. Smith (1941)	51,000
College sophomore	Seashore (1933)	15,000
	Kirkpatrick (1907)	19,000
	Seashore & Eckerson (1940)	60,000
	Gerlach (1917)	85,300
	Gillette (1927)	127,800
	Hartman (1946)	200,000

Note. Adapted from Seashore and Eckerson (1940) and Bayer (1976).

substantial problems of measurement remained. By now, in the time span of educational research, we might want to call the field "ancient," and virtually all of those original problems persist.

There are important practical reasons for attempting to make accurate assessments of total word knowledge. Language and reading programs aim to increase students' vocabularies. The number of words presented to students varies, in part, according to what is regarded as the most authoritative thinking and research on vocabulary size and growth (Clifford, 1978). More reliable estimates would indicate the appropriateness of the assumptions of a program, and perhaps highlight periods of growth to be capitalized upon. More generally, reliable estimates would indicate whether direct language instruction can plausibly account for a substantial proportion of the child's language growth, or whether word knowledge is acquired for the most part independently of formal instruction. To refer again to a concrete proposal, Becker's (1977) idea that underachieving children should be taught via direct instruction the vocabulary most high school seniors possess would be difficult, but perhaps feasible, if the children had to learn 25 new words a week. It would be out of the question if they had to learn 25 words each school day.

Next we will present some of the central issues in broad-gauged measurement of word knowledge. The discussion of these issues will reveal many of the reasons why estimates of vocabulary size have fluctuated so widely. Two general questions need to to be considered. First, how is a sample of words to be selected? Second, what kind of response from a subject will be regarded as evidence that a word is in the individual's vocabulary?

Selecting a Sample of Words

In determining what is to count as a word, the researcher needs to decide whether or not it is of interest to discern the subject's ability to use derivatives and compounds (plurals, participles, tense markers, comparatives, etc.). Some authors, notably Seashore (1933), have preferred to calculate separate estimates for "special" terms and derivatives. Others, for example Dupuy (1974), have attempted to concentrate solely on "basic" words. Dupuy, the author of one of the most recent and thorough studies of word knowledge, sampled randomly from *Webster's Third New International Dictionary* (1961) and then applied three criteria to each word selected. The word had to be a main entry; a single word form (i.e., not a derivative or compound); and could not be technical, slang, foreign, or archaic.

The systematic nature of this sampling creates its own equally systematic biases. Some children may have acquired the generative rule for, say, negation by prefix (for example, *unable* or *dishonest*) and others may not have (Silvestri & Silvestri, 1977). Do we wish to exclude this element of vocabulary knowledge from the measure? Adults acquire a number of special or technical terms in their areas of expertise or interest, so exclusion of technical terms denies many subjects the opportunity of indicating their knowledge of a large number of words.

What counts as a word will depend upon the researcher's principal purposes. However, affixes and derivatives are important elements of word knowledge, and several questions related to their role are of considerable interest: In what way does knowledge of basic or root word forms relate to knowledge of the compound forms? Are entries organized conceptually in the personal dictionary such that the probability of knowing a compound word is the same as that of knowing all its family members, basic form included? Or is the chance of knowing a compound some combination of the frequencies of the particular compounding elements? Much is to be gained from research into these issues.

Whatever criteria are applied, there can be no doubt that there are many thousands of words in English. Dupuy (1974) estimated that there are about a quarter of a million main entries in *Webster's Dictionary* (1961). Of these, he calculated that about 12,300 are basic words.

A source and method of selecting from that source is required which will lead to the most accurate estimates of total word knowledge. The most obvious way to start is to sample randomly from an unabridged dictionary. Dupuy (1974), for instance, selected one word from every page of the dictionary (the third word from the top of alternating columns), and then applied the three criteria mentioned earlier for selecting the basic words out of this group. This procedure produced a final sample of 123 basic words.

Once a random sample of words has been selected, a test is constructed to assess how many of the words a person knows. Then, in principle, estimating the person's vocabulary size is straightforward. For instance, Dupuy's Basic Word Vocabulary Test contains 1 percent of the 12,300 basic words he calculated are in *Webster's*. Therefore, the absolute size of the basic word vocabulary can be approximated by multiplying the score on this test by 100. A person whose score is 60, after correction for guessing, would be judged to have a basic vocabulary of 6,000 words.

One disadvantage of this method is self-evident. Estimated vocabulary size depends heavily on the size of the dictionary. With respect to Dupuy, while he sampled initially from a large unabridged dictionary, a word had to appear as a major entry in each of three other smaller dictionaries in order to be counted as a basic word. A total of 979 words, 41 percent of the sample, were discarded on the basis of this rule. The result was a very conservative estimate of the number of basic words in American English and is one reason Dupuy's estimates of basic vocabulary size are so much smaller than those of other investigators. Of course, many of these words were very rare, but others such as *cloudlet, escaping, breezes, invited, starling*, and *unilateral* would be familiar to most people.

Already discussed is the issue of what to do with derivative and compound forms. A liberal policy will lead to large estimates of vocabulary size. A conservative policy will produce smaller ones. Dupuy was conservative. He eliminated 7.7 percent of the words in his sample on the grounds that they were compounds or derivatives, including a great many familiar ones, such as *grandchild, package*, and *toothache*.

There are other, more subtle considerations in selecting a random sample of words from a dictionary. Some procedures for sampling from an unabridged dictionary can introduce systematic error since all entries do not occupy the same amount of space on a page. This disproportion typically favors the words in more common use since these are the most elaborated, particularly in an unabridged dictionary where very many derivatives may be listed (Williams, 1932; Lorge & Chall, 1963). Consequently, while the words may seem to have been randomly selected, the frequency distribution of the sample may be substantially different from that of the population. This may partly account for the very large estimates of Seashore and Eckerson (1940) and Smith (1941).

A further problem is that projecting a vocabulary size from performance on a random sampling of words is inefficient. If the subject provides the meaning of *bibulous*, then using up test time by asking for the meaning of *bicycle* is wasteful. When estimating subjects' total vocabulary size is the researcher's major aim, then efficiency of items covered per unit of examinee time is an important consideration.

One obvious response to these problems is to select the sample from a frequency distribution of words. Terman and Merrill (1937) arranged their sample of words in order of "difficulty." When the subject failed at six consecutive words, the vocabulary test was stopped. Dupuy (1974) recommended a similar procedure. Time can be saved by such a procedure, but vocabulary size is likely to be underestimated. Furthermore, heavy stress is placed on the assumption that the frequency distribution of the sample mimics that of the population. If this assumption fails, then multiplication of the subject's score by the appropriate constant will produce a poor estimate of total words known.

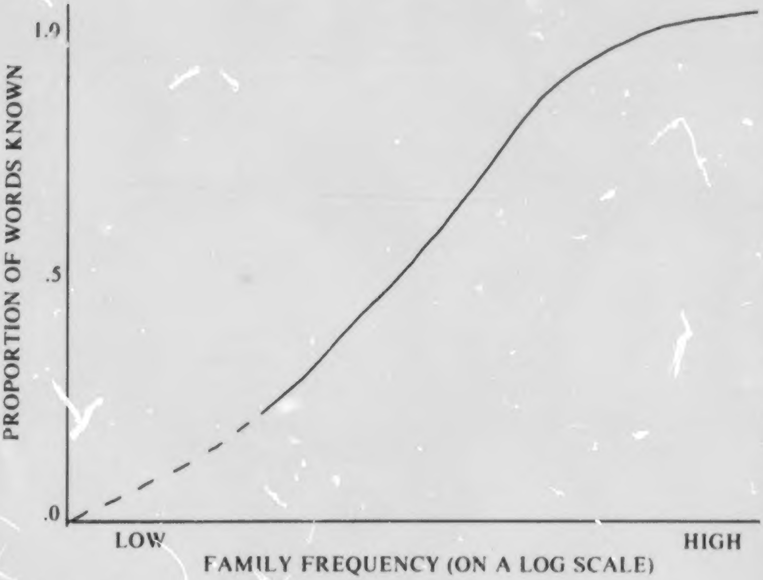
The characteristics of the two major current word frequency compilations available (Carroll, Davies, & Richman, 1971; Kučera & Francis, 1967) suggest a potential problem with frequency sampling. These analyses indicate that the distribution of words is highly unbalanced, a conclusion reached over 25 years ago by Horn (1954), who calculated that about 2,000 types will account for about 95 percent of "running words in adult writing"; 3,000 for 96.9 percent; 4,000 for 97.8 percent; and 10,000 for 99.4 percent. At the low frequency end of the scale, there is a tail that approaches infinity. Even in a huge corpus, a vast number of words appear only once, twice, or not at all. Of the 86,741 word types listed by Carroll, Davies, and Richman from a corpus of over 5 million tokens, 35,079, or 40.44 percent, appeared once. Kucera and Francis found 44.72 percent of the words appeared once in a sample of over one million tokens. So, if the test is short, the subjects run the risk of not being able to show that they know several medium frequency words, since there will be such a large proportion of rare words in the sample. A resolution of this issue is important, since a frequency-based sampling technique seems the most accessible method for overcoming the problems of simple random sampling.

Frequency is a parameter which probably is very strongly related to probability that a word will be known. There is evidence supporting this hypothesis from a number of areas: multiple choice performance on standardized tests (Kibby, 1977), recall of word meanings following presentation of pictures (Carrol & White, 1973; Duncan, 1977), and word recognition times following tachistoscopic presentation (Cohen, 1976; Rubenstein, Garfield, & Millikan, 1970). The only discrepant finding has been that of Davis (1944b) who found only a slight relationship between word difficulty and frequency. He explained this result in terms of the role of compound words: While the root of the word may be very common and well-known, a certain affix-root compound may be very infrequent but almost equally well-known if the affix is familiar. A more analytic approach to the relationship of this index of frequency of usage to probability of knowledge would entail the use of "family" frequency, that is, the frequency of the root word and all its compounds and derivatives. We might expect that the relationship of this index of frequency of usage to probability of knowledge would be more orderly.

Indeed, we are willing to go further and speculate that the relationship between family frequency and probability of knowing a word resembles the curve presented in Figure 1. In terms of breadth of knowledge, we would expect a ceiling at the upper end of the frequency scale: most people know all of the very common words. Other aspects of the curve would differentiate individuals: The point at which the curve dropped from the plateau level, and the slope of the function probably are the two parameters that would capture the important individual differences. Even for children, we might best think of the curve leveling out as the words become very infrequent, since it is likely that, from their hobbies, interests, or the occupation of their parents, most children would know some very rare words. Nevertheless, we have drawn the lower portion of the curve as a broken line since we are less sure about the relationship in this area.

In summary, a good test of word knowledge would present the subject with a large number of words, sampled liberally from the whole range of word frequency. Techniques

Figure 1.
Possible relationship between likelihood word meanings are known and frequency of usage.



should be developed which allow accurate estimation of the relationship of a given subject's probability of knowing a word and the frequency of the word's morphological family.

Criteria for Determining that a Word Is in a Person's Vocabulary

Four sorts of test formats have been employed in attempts to assess breadth of vocabulary knowledge: a) multiple choice; b) constructed answer in which the subject attempts to give a definition, a synonym, an illustration, or use the word in a sentence or phrase; c) yes/no judgments, in which the subject checks the words in a list that he or she knows; and d) matching where the subject pairs off words with their synonyms. Sims (1929) compared these four types using data obtained from students in fifth through the eighth grades. The correlation matrix Sims reported is reproduced in Table 4. Sims concluded that, although the checking method was as reliable as the others, it did not seem to offer acceptable construct validity. Only seventy words were used, however, and Sims failed to counterbalance for order or delay between tests. While there may be some questions about the trustworthiness of Sims' results, there is intuitive sense in the notion that the constructed answers, multiple choice, and matching tasks have more in common with one another than they have with a checking task that is not corrected for guessing.

Table 4
Correlations between Four Types of Vocabulary Tests

	1	2	3	4
1. Checking (yes/no)	.92*			
2. Multiple choice	.54	.84*		
3. Matching	.64	.85	.93*	
4. Constructed answer	.56	.74	.82	.92*

Note. From Sims (1929).
*Split-half reliability coefficients.

The question that needs examination is which of these methods will be of most theoretical and practical value as a measure of vocabulary. Three of these types will be discussed in the light of several issues. Since the points raised about the multiple choice format apply even more cogently to matching, the latter will not be dealt with separately.

Multiple choice methods. People often possess partial knowledge of words. In these instances the items' distractors become crucial. An individual may select the correct synonym for *platitude* from the choices: a) duck-billed mammal, b) praise, c) commonplace remark, d) flatness. He may make the correct selection because he has heard the word used in reference to an utterance and with a negative connotation. This information, however, may not enable him to select correctly from a) commonplace remark, b) nonsense, c) irrelevant question, d) insult. The set of choices constrains the individual's response to different degrees, and different policies for generating distractors will, of course, lead to differences in performance.

Lepley (1955, 1965), for example, constructed two forms of a synonym test, one employing distractors from the same semantic category as the target, and another which used distractors from semantically diverse categories. Lepley (1965) found equal split-half reliability (.93 and 94) but only a .66 correlation between performance on the two scales, and significantly superior performance on the version requiring only gross discriminations. The correlation is surprisingly low given the common format and the fact that the superficial demand characteristics were the same. Lepley's results illustrate the influence of the distractor set.

The multiple choice format is currently the most widely used in standarized vocabulary testing (e.g., California Achievement Tests, 1977; Metropolitan Achievement Tests, 1970; Stanford Achievement Tests, 1973). The principal complaint raised here so far is that the distractors cannot avoid constraining the subject's response. If the purpose of the test is to provide data on relative performance only, not on absolute level of performance, then the distractors can be, and usually are, chosen to maximize the discriminating power of the item. If one is interested in vocabulary size, then this policy will not do.

Many vocabulary tests (e.g., Stanford, 1973) use sentence completion in a multiple choice format. Many of the problems already mentioned apply even when the test simulates a real encounter with the target word. In addition, the question of the effects of various amounts of contextual support on estimated vocabulary size, with groups of words that vary in frequency of usage, has not been studied. There is research that suggests that individuals vary not only in the size of their reading vocabularies but also in their ability to use context to deduce the meanings of unknown and partly known words (Mason, Knisely, & Kendall, 1978; Pearson & Studt, 1975).

A tricky problem with the multiple choice format is that young children may not consider all the distractors (Asher, 1978; Brown, 1975; Vurpillot, 1968). They will often choose the first or second alternative if it makes reasonable enough sense. The test-taking strategies of older children on multiple choice tests are not yet well characterized, but there quite probably are strategic components of good performance which serve to increase spuriously the relationship between a multiple choice vocabulary test and other achievement or intelligence tests in the same format. An insidious possibility is that some of the apparent growth in vocabulary knowledge over the elementary school years is really attributable to the acquisition of more sophisticated test-taking skills.

In conclusion, the multiple choice format is the most popular one. It makes relatively efficient use of examinee time and must be reasonably valid, otherwise the strong relationships between performance on such tests and other measures of linguistic competence, summarized at the beginning of this paper, would not have been obtained. The chief complication with the multiple choice format, when one wants absolute measures of vocabulary knowledge, is how to choose distractors. A further problem is that multiple choice tests may make demands on strategic knowledge in which young and poor readers are deficient.

Constructed answer measures. To overcome the problem of selecting distractors, several researchers, notably Seashore (1933), Smith (1941), Terman and Merrill (1937), have used a constructed answer format in which the subject reads or hears the target word and then writes or tells a definition of it, uses it in a sentence, gives a synonym for it, or in some other way provides an indication of its sense and reference. Subjects can be encouraged to do any one of these things just so long as the experimenter is convinced the word is "known." This format is capable of dealing with a variety of levels of knowing a word and avoids the issue of distractors. There are, however, two substantial problems with constructed answer measures: The problem of scoring the answers and the problem of response bias.

In the written format, in particular, a constructed answer measure is confounded by factors such as spelling ability, sentence construction ability, and even the ability to write legibly, all of which may discourage a subject from elaborating on a word used or understood in conversation. A slightly more subtle problem, and one that is more difficult to control, resides in the fact that, if a liberal criterion is used and the subject is allowed a range of possible responses to a target word, then a particular strategy for responding may be adopted. The problem is that some words would be more easily explicated in a particular form. The word *noun* may be more easily explained through illustration than by definition, for instance. The research of Anglin (1970) and Wolman and Baker (1965) indicates that, up to the age of about 10-12 years, children tend to provide concrete definitions-by-illustration rather than by an inclusive term or synonym. It is entirely possible that, depending on scoring criteria, the preference at a different age for certain explanatory strategies could produce spurious estimates of the rate of vocabulary growth.

A really vexing problem is how liberally to score answers. How does one score synonyms in relation to apt illustrations or perfect usage in a sentence? In many instances, partial knowledge is displayed. In one of our own recent testing sessions, it became clear that many fifth grade students had partial knowledge of the word *forbid*. Several students knew that it had something to do with not being permitted to do something but did not have as part of their knowledge the fact that *forbid* is used in imperative speech acts. We soon realized that, in this case, we needed to ask for its use in a sentence. We have found other more subtle and difficult cases of partial knowledge. For the word *propelled*, there was no problem in the students' recognition of the word because of their knowledge of *propeller*. When probed about the function of a propeller, many came close to generating the notion of propulsion on the theory that it would be strange to have a big round blade going around on the front of a plane unless it served some fairly fundamental purpose—and what planes do is move.

Some words have no near-synonyms. There are other instances when the only synonym is a less frequent word than the target. In such cases, the subject is being asked to produce a rare word in order to show that a common word is known.

There are some almost irresistible tendencies displayed by an examiner when administering a test with a constructed answer format. After a few children have been tested, the examiner develops a sense of which words are easy and which are difficult. It requires conscious effort to avoid expecting more explanation of the difficult words and less for the easy words. If every subject has known *chair* and the current subject pats the seat of his stool as a response, then the tendency is to award full marks. If he pats the wall for *edifice*, however, he might not score so well. Similarly there is an urge to expect more elaborated responses from older subjects. The preschooler who tells you that an *automobile* "goes brrrrrmmm" will strike you more favorably than the college sophomore who gives you the same answer. In addition, the experimenter will witness explanations of words which entail subtle nonverbal as well as verbal cues. Young children typically employ hand movements, facial expressions, and gesture in their communications especially when dealing with words that are a little difficult for them.

The horns of the dilemma are these. Stringent, operational, adult-like standards for evaluating whether a response indicates a word is known will confound what is supposed to be a measure of breadth of vocabulary knowledge with expository ability. Looser, more flexible standards will confound the measure with the subjective judgment of the examiners which may change from word to word, subject to subject, and occasion to occasion.

So the liabilities of the constructed answer method are both logistical and substantial. It is inefficient per unit of testing and scoring time, and it seems to rely on often subtle intuitions on the part of the examiner, especially when the subject displays partial knowledge of an item.

Yes/no format. The final format to be considered is that of "checking," which we prefer to term a yes/no method. In this format the subject simply indicates whether or not the meaning of a word is known. Two of the major difficulties that have arisen consistently in the discussion of the other two major formats are the problem of response bias, and the need to present the subject with a large number of words chosen from a wide frequency range. The checking format can satisfy the second criterion admirably but problems of validity arise. Sims (1929) concluded:

> The writer is inclined to believe that a good guess as to whether or not a child knows the meaning of a word is almost as satisfactory a method of determining vocabulary as checking tests. The relative simplicity of such a measure, the ease of preparation and administration should not blind one to its invalidity. (p. 96)

Chall and Dale (1950) reported that the average tendency to overestimate word knowledge in the yes/no format over and above the definition format amounted to about 11 percent, and was more pronounced for rare words.

It ought to be no real surprise that a yes/no test uncorrected for decisions in the face of partial knowledge would give inflated estimates of vocabulary size and would correlate poorly with other measures. Consider the yes/no task from the point of view of the test taker. Some individuals may deny that they know the word *gold* because they do not know its atomic weight, while others will agree they know it because they have a feeling that it can be used to refer to a color.

The problem of correcting yes/no test scores for guessing is not insuperable. Stating the issue more precisely, guessing is only part of the problem. The real issue, as the *gold* example illustrates, is one of eliminating variation in the degree of confidence different individuals must have before they are willing to say, "Yes, I know that word."

Signal detection theory (Swets, 1964) affords a conceptual and computational framework that may allow estimation of amount of word knowledge independent of judgmental standards. This theory was originally developed for use in psychophysical experimentation. In this setting, typically the subject is informed that he will hear a short burst of background noise and that there may be a tone sounded as well. The subject's task is to report whether or not a tone (the signal) was present. Research has established that it is possible to get a very accurate estimate of a person's capacity to detect the signal by correcting for whatever tendency he or she has to report "hearing" the signal when it is not actually there. Pastore and Scheirer (1974) have summarized research showing that this paradigm can be applied to the analysis of a broad range of perceptual and cognitive tasks. With respect to vocabulary assessment, the work of Zimmerman, Broder, Shaughnessy, and Underwood (1977) has suggested that, by using close-to-English nonsense letter strings as the "noise only" stimuli, signal detection methods might be applied to word knowledge.

We are currently analyzing data collected from elementary and high school subjects on large numbers of words. The students responded yes or no to a mixture of many English words and almost as many nonsense words. Later they completed standardized multiple choice questions on the real words. Our preliminary analyses have indicated that yes/no scores adjusted according to signal detection theory, and other corrections for guessing and risk-taking, correlate highly with multiple choice performance. We later interviewed the subjects individually about a subset of the words. The data suggest that a value derived from the yes/no task gives a better estimate of true word knowledge than performance on the standardized multiple choice test.

The fact that words have multiple meanings poses a problem for the yes/no task, since presumably a person will check "yes" if he or she knows *any* meaning of a word. This is not a small problem. According to Lovell (1941), 43 percent of the words used by Seashore and Eckerson (1940) had multiple meanings. Recently, Balch (cited in Johnson & Pearson, 1978, p. 17) has reported that from 23 percent to 42 percent of the words in six widely used basic vocabulary lists have multiple meanings. In other recent research, Mason, Knisely, and Kendall (1978) have shown that children are much less likely to know the secondary than the primary meaning of words used in their secondary sense in a popular basal series. It is apparent that the yes/no format is not suitable for distinguishing which of the meanings of a word are known. When that is the goal, some other method of assessment is required.

In summary, the great attraction of the yes/no format is that it permits the presentation of a very large number of words in a given interval of examinee time. Compared to the multiple choice format, it reduces somewhat the burden of preparing distractors and, compared to constructed answer formats, it sidesteps vagaries of scoring. The notable problem with the yes/no task is that scores of individuals will be influenced markedly by differences in tendency to take risks in the face of uncertainty. If this problem can be solved, the yes/no task might be very useful for assessment of breadth of word knowledge.

Conclusion

While current research demonstrates the importance of such factors as a reader's perspective on a text (Pichert & Anderson, 1977) and text structure (Mandler & Johnson, 1977; Meyer, 1975), it is also clear that word knowledge is a requisite for reading comprehension: people who do not know the meanings of very many words are most probably poor readers. There are serious gaps in our understanding of why this is true and of how word knowledge grows throughout the life span. Filling those gaps promises to be both an intellectual and a practical challenge of considerable importance. We judge that a critical first step is the development of improved methods of assessing breadth of vocabulary knowledge. It is only after some refinement has been achieved at this level that models of lexical development and instructional programs can be based on realistic expectations about the acquisition of word meanings.

We conclude our review of vocabulary knowledge and vocabulary size with the realization that, since the turn of the century, a tremendous amount of energy has been put into answering the question, "How many words does an individual know?" We have come to wonder if this question is properly framed. The nature of language may make it unanswerable and thus, for scientific purposes, irrelevant. Empirical methods may be able to generate useful indices such as that discussed earlier—the relationship of the individual's knowledge of words to word frequency. However, to produce a single value from performance on a sample to represent total vocabulary size may be an exercise that relies too heavily on the assumptions of a static population of isolated words and on an overly restrictive view of how we generate and use words in context.

References

ALLEN, L.R., & GARTON, R.F. The influence of word knowledge on the word frequency effect in recognition memory. *Psychonomic Science*, 1968, *10*, 401-402.

AMES, W.S. The understanding vocabulary of first grade pupils. *Elementary English*, 1964, *41*, 64-68.

ANDERSON, R.C. Schema-directed processes in language comprehension. In A.M. Lesgold, J.W. Pellegrino, S.D. Fokkema, & R. Glaser (Eds.), *Cognitive psychology and instruction*. New York: Plenum, 1978.

ANDERSON, R.C., & ORTONY, A. On putting apples into bottles: A problem of polysemy. *Cognitive Psychology*, 1975, *7*, 167-180.

ANDERSON, R.C., PICHERT, J.W., GOETZ, E.T., SCHALLERT, D.L., STEVENS, K.W., & TROLLIP, S.R. Instantiation of general terms. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 667-679.

ANGLIN, J.M. *The growth of word meaning*. Cambridge, Mass.: M.I.T. Press, 1970.

ASHER, S.R. *Referential communication* (Tech. Rep. No. 90). Urbana: University of Illinois, Center for the Study of Reading, June 1978.

BAYER, M.L. *Primary grade understanding vocabulary as measured by an orally administered Basic Word Vocabulary Test*. Unpublished master's thesis, Rutgers University, 1976. (ERIC Document Reproduction Service No. ED 131 417)

BECKER, W.C. Teaching reading and language to the disadvantaged—what we have learned from field research. *Harvard Educational Review*, 1977, *47*, 518-543.

BONSER, F.G., BURCH, L.H, & TURNER, M.R. Vocabulary tests as measures of school efficiency. *School and Society*, 1915, *2*, 714-718.

BORMUTH, J.R. Readability: A new approach. *Reading Research Quarterly*, 1966, *1*, 79-132.

BOTZUM, W.A. A factorial study of reasoning and closure factors. *Psychometrika*, 1951, *16*, 361-386.

BRANDENBURG, G.C. Psychological aspects of language. *Journal of Educational Psychology*, 1918, *9*, 313-332.

BREWER, W.F., & STONE, B.J. Acquisition of spatial antonym pairs. *Journal of Experimental Child Psychology*, 1975, *19*, 199-307.

BROWN, A.L. Recognition, reconstruction and recall of narrative sequences by preoperational children. *Child Development*, 1975, *46*, 156-166.

California Achievement Tests. Monterey, Calif.: CTB/McGraw-Hill, 1977.

CARROL, J., & WHITE, M. Age-of-acquisition nouns for 200 picturable nouns. *Journal of Verbal Learning and Verbal Behavior*, 1973, *12*, 563-576.

CARROLL, J.B., DAVIES, P., & RICHMAN, B. *Word frequency book*. Boston: Houghton Mifflin, 1971.

CHALL, J.S. *Readability: An appraisal of research and application*. Columbus: Ohio State Press, 1958.

CHALL, J.S., & DALE, E. Familiarity of selected health terms. *Educational Research Bulletin*, 1950, *39*, 197-206.

CLARK, E.V. What's in a word? On the child's acquisition of semantics in his first language. In T.E. Moore (Ed.), *Cognitive development and the acquisition of language*. New York: Academic Press, 1973.

CLARK, H.H., & CLARK, E.V. *Psychology and language*. New York: Harcourt Brace Jovanovich, 1977.

CLARK, N.L. *Hierarchical structure of comprehension skills*. (2 vols.). Hawthorn, Vict., Austral.: A.C.E.R, 1972.

CLIFFORD, G.J. Words for school: The applications in education of the vocabulary researches of E.L. Thorndike. In P. Suppes (Ed.), *Impact of research on education: Some case studies*. Washington, D.C.: National Academy of Education, 1978.

COHEN, J. Tachistoscopic word recognition as a function of word abstractness/concreteness, word frequency, and IQ. *Perceptual and Motor Skills*, 1976, *42*, 471-476.

COLEMAN, E.B. Developing a technology of written instruction: Some determiners of the complexity of prose. In E.Z. Rothkopf & P.E. Johnson (Eds.), *Verbal learning research and the technology of written instruction*. New York: Columbia University, Teachers College Press, 1971.

CUFF, N.B. Vocabulary tests. *Journal of Educational Psychology*, 1930, *21*, 212-220.

DAVIS, F.B. Fundamental factors of comprehension in reading. *Psychometrika*, 1944, *9*, 185-197. (a)

DAVIS, F.B. The interpretation of frequncy ratings obtained from *The Teacher's Wordbook*. *Journal of Educational Psychology*, 1944, *35*, 169-174. (b)

DAVIS, F.B. Research in comprehension in reading. *Reading Research Quarterly*, 1968, *3*, 499-545.

DOLCH, E. How much word knowledge do children bring to grade 1? *Elementary English Review*, 1936, *13*, 177-183.

DUNCAN, C.P. Effects of frequency on retrieval from a semantic category, *Bulletin Psychonomic Society*, 1977, *10*, 57-59.

DUPUY, H.P. *The rationale, development and standardization of a basic word vocabulary test*. Washington, D.C.: U.S. Government Printing Office, 1974. (DHEW Publication No. HRA 74-1334)

ELWOOD, M.I. A preliminary note on the vocabulary test in the revised Stanford-Binet Scale. *Journal of Educational Psychology*, 1939, *30*, 632-634.

FILLMORE, C.J. The future of semantics. In R. Austerlitz (Ed.), *The scope of American linguistics*. Amherst, Mass.: Ridder Press, 1975.

FLEISHER, L.S., & JENKINS, J.R. *Effects of contextualized and discontextualized practice conditions on word recognition* (Tech. Rep. No. 54). Urbana: University of Illinois, Center for the Study of Reading, July 1977.

FRUCHTER, B. The nature of verbal fluency. *Educational and Psychological Measurement*, 1948, *8*, 33-47.

GENTNER, D. Evidence for the psychological reality of semantic components: The verbs of possession. In D.A. Norman, D.E. Rumelhart, & the LNR research group (Eds.), *Explorations in cognition*. San Francisco: Freeman, 1975.

GERLACH, F.M. *Vocabulary studies - false definition test*. Colorado: College Studies in Education and Psychology, No. 1, 1917.

GILLETTE, J.M. Extent of personal vocabularies and cultural control. *Science Monthly*, 1927, *29*, 451-467.

GLUCKSBURG, S., HAY, A., & DANKS, J.H. Words in utterance contexts: Young children do not confuse the meaning of *same* and *different*. *Child Development*, 1976, *47*, 737-741.

GOODMAN, K.S. Behind the eye: What happens in reading. In H. Singer & R.B. Ruddell (Eds.), *Theoretical models and processes of reading*. Newark, Del.: International Reading Association, 1976.

HALFF, H.M., ORTONY, A., & ANDERSON, R.C. A context-sensitive representation of word meanings. *Memory and Cognition*, 1976, *4* (4), 378-383.

HALL, W.S., & TIRRE, W.C. *The communicative environment of young children: Social class, ethnic and situational differences* (Tech. Rep. No. 125). Urbana: University of Illinois, Center for the Study of Reading, May 1979.

HARTMAN, G.W. Further evidence on the unexpected large size of recognition vocabularies among college students. *Journal of Educational Psychology*, 1946, *37*, 436-439.

HOLLEY, C.E. *Holley sentence vocabulary scale - grades 3-12*. Bloomington, Ill.: Public School Publication Co., 1919.

HORN, E. *Teaching spelling*. Washington, D.C.: A.E.R.A., National Education Association, 1954.

HUNT, E. Mechanics of verbal ability. *Psychological Review*, 1978, *85*, 109-130.

HUNT, E., LUNNEBORG, C., & LEWIS, J. What does it mean to be high verbal? *Cognitive Psychology*, 1975, *7*, 194-227.

JENKINS, J.R., PANY, O., & SCHRECK, J. *Vocabulary and reading compre-*

hension: Instructional effects (Tech. Rep. No. 100). Urbana: University of Illinois, Center for the Study of Reading, August 1978.

JOHNSON, D.D., & PEARSON, P.D. *Teaching reading vocabulary*. New York: Holt, Rinehart & Winston, 1978.

KIBBY, M.W. A note on the relationship of word difficulty and word frequency. *Psychological Reports*, 1977, *41*, 12-14.

KIRBY, J.R., & DAS, J.P. Reading achievement, IQ, and simultaneous-successive processing. *Journal of Educational Psychology*, 1977, *69*, 564-570.

KIRKPATRICK, E.A. Vocabulary test. *Popular Science Monthly*, 1907, *70*, 157-164.

KLARE, G.R. Assessing readability. *Reading Research Quarterly*, 1974-1975, *10*, 62-102.

KUČERA, H., & FRANCIS, W.N. *Computational analysis of present-day American English*. Providence, R.I.: Brown University Press, 1967.

LaBERGE, D., & SAMUELS, S.J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293-323.

LABOV, W. The boundaries of words and their meanings. In C.J. Bailey & R. Shuy (Eds.), *New ways of analyzing variation in English*. Washington, D.C.: Georgetown University Press, 1973.

LEPLEY, W.M. Rationale, construction and preliminary try-out of the synonym vocabulary test. *Journal of Psychology*, 1955, *39*, 215-225.

LEPLEY, W.M. The synonym vocabulary test: Form III. *Journal of Psychology*, 1965, *59*, 109-112.

LEWINSKI, R.J. Vocabulary and mental measurement: A quantitative investigation and review of research. *Journal of Genetic Psychology*, 1948, *72*, 247-281.

LORGE, I., & CHALL, J. Estimating the size of vocabularies of children and adults: An analysis of methodological issues. *Journal of Experimental Education*, 1963, *32*, 147-157.

LOVELL, G.D. Interrelations of vocabulary skills: Commonest versus multiple meanings. *Journal of Educational Psychology*, 1941, *21*, 67-72.

MAHAN, H.C., & WITMER, L. A note on the Stanford-Binet vocabulary test. *Journal of Applied Psychology*, 1936, *20*, 258-263.

MANDLER, J.M., & JOHNSON, N.S. Remembrance of things parsed: Story structure and recall. *Cognitive Psychology*, 1977, *9*, 11-151.

MARKS, C.B., DOCTOROW, M.J., & WITTROCK, N.C. Word frequency and reading comprehension. *Journal of Educational Research*, 1974, *67*, 259-262.

MASON, J.M., KNISELY, E., & KENDALL, J. *Effects of polysemous words on* *sentence comprehension* (Tech. Rep. No. 85). Urbana: University of Illinois, Center for the Study of Reading, May 1978.

McNEMAR, Q. *The revision of the Stanford-Binet Scale.* Boston: Houghton-Mifflin, 1942.

Metropolitan Achievement Tests. New York: Harcourt Brace Jovanovich, 1970.

MEYER, B.J.F. *The organization of prose and its effects on memory.* Amsterdam: North Holland, 1975.

MILLER, G.A. The acquisition of word meaning. *Child Development*, 1978, *49*, 999-1004.

MILLER, G.A., & JOHNSON-LAIRD, P.N. *Language and perception.* Cambridge, Mass.: Belknap Press, 1976.

MINER, J.B. *Intelligence in the U.S.* New York: Springer, 1957.

NELSON, K. Cognitive development and the acquisition of concepts. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge.* Hillsdale, N.J.: Erlbaum, 1977.

PANY, D., & JENKINS, J.R. *Learning word meanings: A comparison of instructional procedures and effects on measures of reading comprehension with learning disabled students* (Tech. Rep. No. 25). Urbana: University of Illinois, Center for the Study of Reading, March 1977.

PASTORE, R.E., & SCHEIRER, C.J. Signal detection theory: Considerations for general application. *Psychology Bulletin*, 1974, *81*, 945-958.

PEARSON, P.D., & STUDT, A. Effects of word frequency and contextual richness on children's word identification abilities. *Journal of Educational Psychology*, 1975, *67*, 89-95.

PERFETTI, C.A., & LESGOLD, A.M. Coding and comprehension in skilled reading and implications for reading instruction. In L.B. Resnick & P.A. Weaver (Eds.), *Theory and practice of early reading* (Vol. 1). Hillsdale, N.J.: Erlbaum, 1979.

PICHERT, J.W., & ANDERSON, R.C. Taking different perspectives on a story. *Journal of Educational Psychology*, 1977, *69*, 309-315.

POSNER, M.I., & MITCHELL, R. Chronometric analysis of classification. *Psychology Review*, 1967, *74*, 392-409.

PUTNAM, H. The meaning of 'meaning.' In H. Putnam (Ed.), *Mind, language and reality.* New York: Cambridge University Press, 1975.

RAVEN, J.C. The comparative assessment of intellectual ability. *British Journal of Psychology*, 1948, *39*, 12-19.

RICHARDS, M.R. *Come* and *go* reconsidered: Children's use of deictic verbs in contrived situations. *Journal of Verbal Learning and Verbal Behavior*, 1976, *15*, 655-665.

ROSCH, E. On the internal structure of perceptual and semantic categories. In T.E. Moore (Ed.), *Cognitive development and the* *acquisition of languages*. New York: Academic Press, 1973.

ROSCH, E., & MERVIS, C.B. Family resemblances: Studies in the internal structure of categories. *Cognitive Psychology*, 1975, *1*, 573-605.

RUBENSTEIN, H., GARFIELD, L., & MILLIKAN, J. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behavior*, 1970, *9*, 487-494.

SEASHORE, R.H. Measurement and analysis of extent of vocabulary. *Psychological Bulletin*, 1933, *30*, 709-710.

SEASHORE, R.H., & ECKERSON, L.D. The measurement of individual differences in general English vocabularies. *Journal of Educational Psychology*, 1940, *31*, 14-38.

SHIBLES, B.H. How many words does a first grade child know? *Elementary English*, 1959, *31*, 42-47.

SILVESTRI, S., & SILVESTRI, R. Developmental analysis of the acquisition of compound words. *Language, speech and hearing services in the school*, 1977, *8*, 217-221.

SIMS, V.M. The reliability and validity of four types of vocabulary test. *Journal of Educational Research*, 1929, *20*, 91-96.

SMITH, M.E. An investigation of the development of the sentence and the extent of vocabulary in young children. *University of Iowa studies in child welfare*, 1926, *3*, 92.

SMITH, M.K. Measurement of the size of general English vocabulary through the elementary grades and high school. *General Psychological Monographs*, 1941, *24*, 311-345.

SPACHE, G. The vocabulary tests of the revised Stanford-Binet as independent measures of intelligence. *Journal of Educational Research*, 1943, *36*, 512-516.

Stanford Achievement Tests. New York: Harcourt Brace Jovanovich, 1973.

STEFFENSEN, M.S., JOAG-DEV C., & ANDERSON, R.C. A cross-cultural perspective on reading comprehension. *Reading Research Quarterly*, 1979, *15* (1), 10-29.

SWETS, J.A. (Ed.). *Signal detection and recognition by human observers*. New York: Wiley, 1964.

TERMAN, L.M. *The measurement of intelligence*. Boston: Houghton-Mifflin, 1916.

TERMAN, L.M. Vocabulary test as a measure of intelligence. *Journal of Educational Psychology*, 1918, *9*, 452-466.

TERMAN, L.M., & MERRILL, M.A. *Measuring intelligence*. Boston: Houghton-Mifflin, 1937.

THORNDIKE, R.L. *Reading comprehension education in fifteen countries*. New York: Wiley, 1973.

THURSTONE, L.L. A note on a reanalysis of Davis' reading tests. *Psychometrika*, 1946, *11*, 185-188.

TUINMAN, J.J., & BRADY, M.E. How does vocabulary account for variance on reading comprehension tests? A preliminary instructional analysis. In P. Nacke (Ed.), *23rd N.R.C. Yearbook*. Clemson, S.C.: The National Reading Conference, 1974.

VURPILLOT, E. Judging visual similarity: The development of scanning strategies and their relation to differentiation. *Journal of Experimental Child Psychology*, 1968, *6*, 632-650.

WECHSLER, D. *Manual, Wechsler intelligence scale for children*. New York: Psychological Corporation, 1949.

WHITE, M.L. Mental age norms for vocabulary scores on the 1937 Stanford-Binet. *Psychological Records*, 1942, *5*, 159-169.

WILLIAMS, H.M. Some problems of sampling in vocabulary tests. *Journal of Experimental Education*, 1932, *1*, 131-133.

WITTGENSTEIN, L. *Philosophical investigations*. New York: Macmillan, 1953.

WITTROCK, N.C., MARKS, C., & DOCTOROW, M. Reading as a generative process. *Journal of Educational Psychology*, 1975, *67*, 484-489.

WOLMAN, R.N., & BAKER, E.N. A developmental study of word definitions. *Journal of General Psychology*, 1965, *107*, 159-166.

WRIGLEY, C., SAUNDERS, D.R., & NEWHAUS, J.O. Application of the quatrimax method of rotation to Thurstone's P.M.A. study. *Psychometrika*, 1958, *23*, 115-170.

ZIMMERMAN, J., BRODER, P.K., SHAUGHNESSY, J.J., & UNDERWOOD, B.J. A recognition test of vocabulary using Signal-Detection Measures, and some correlates of word and nonword recognition. *Intelligence*, 1977, *1*, 5-31.

Social Context of Learning to Read

Courtney B. Cazden
Harvard University

To adults, reading is a solitary activity, a kind of internal language process that contrasts with interpersonal talk. The contrast is not complete: We read song sheets aloud and together; we exchange notes during a lecture, thus using reading as well as writing for immediate interactional ends; and we listen alone to talk on the radio or TV, thus making a solitary activity of the comprehension of speech. But usually we talk with others and read alone.

Not so with children, especially children just learning to read in the primary grades. Learning to read, like mature reading later on, is certainly a cognitive process; but it is also a very social activity, deeply embedded in interactions with teacher and peers. Hopefully, as we understand those interactions more fully, we will be able to design more effective environments for helping children learn. This paper reviews research on children learning to read in classroom interactions in four parts: influences on time engaged in reading; differences in the focus of instruction; the complexities of reading group lessons; and peer interactions in the older grades.

Influences on Time Engaged in Reading

One obvious way in which classroom interactions affect learning is through their effect on how much time children actually spend engaged in reading tasks. Three descriptions by

Piestrup (1973), McDermott (1978), and Au (in press, in preparation), are sociolinguistic and ethnographic analyses of audio- or videotaped lessons. Two descriptions by Hess and Takanishi (1974) and Cazden (1973) are more traditional studies in which observers did on-the-spot coding.

Piestrup's research (1973) is on sources of interference between the language of black children and their teachers. In an analysis of 104 reading instruction episodes audiotaped in 14 first grade classrooms with predominantly black children, Piestrup identified two kinds of interference which she labelled structural and functional. Whether the mismatch is only a temporary misunderstanding or a more serious barrier depends on the teacher's understanding of the problem and her response to it. In the following episode about a workbook lesson, the teacher explicitly and effectively dealt with a structural (dialect) conflict:

T "... how would you harm the colt?"

C_1 Tear it.

T Huh?

C_1 Tear it.

T Th—th—Oh! Do you, do you know what a colt is, now?

C_1 Oh, kill it, kill it!

T No, what's a colt?

C_1 Somethin' you wear.

T There's an "l" in it. "Coat" is c-o-a-th—don't laugh, that's all right. "Colt" is very hard for city children, because they haven't been out on the farm, and they don't know about it. It's a baby, a baby colt.

C_1 A baby colt.

C_1 Oh yeah!

T Remember the story? An' it's a c-o-l-t. "Coat" is c-o-a-t, and there's no "l" in it, but listen to—Keisha—co*l*t, co*l*t, co*l*t. Now do you know what a colt is?

This is an expanded version of a paper by the same title which appears in L.B. Resnick & P.A. Weaver (Eds.), *Theory and Practice of Early Reading*. Reprinted by permission of Lawrence Erlbaum Associates, Inc.

C_4 Yeah, I know.
T What is it?
C_2 A baby horse.
T Yes, uh-huh, how could you harm a baby horse?

(Piestrup, 1973, pp. 3-5)

Interference is termed functional rather than structural when the mismatch comes from the functions language is used for rather than from structural features of the language itself. In the following excerpt from oral reading, the children shift away from discussion of remote content to verbal play; the teacher is ignored and fails to get their attention back to the reading task:

T "Off"
C_1 "Off to the—
T OK. It says "wood."
C_1 —wood.
T We would say woods—this book was written in England.
C_1 Now, I'm through. I ain't gonna read this page again.
T OK. Well, we're gonna turn the page and we're just gonna read the next page.
C_1 Uh uh! Darren 'sposed to be first.
T Well, I'm waiting for Darren to come back. Come on, Darren.
C_2 He just playin' aroun' ______________ (not clear).
C_1 He crack his knuckles, in the buckles.
C_3 Uh-uh.
T OK, Zip and Wendy ran to the woods, and here's the —
C_1 I got a tow truck. My mama bought me one.
T —father.
C_1 An' I got me a car to hook it on. It got a hook—

(Piestrup, 1973, pp. 6-7)

The two teachers out of the group of fourteen who were able to accommodate most effectively to both structural and functional sources of interference, termed "Black Artful" by Piestrup, had teaching episodes that were both lively and focused on reading, and their children had the highest reading scores at the end of first grade. Piestrup concludes that "the ways teachers communicate in the classroom are crucial to children's success in learning to read" (p. 170).

McDermott (1978) has done an intensive microanalysis, frame by frame, of videotapes of two 30-minute reading groups (top group and bottom group) in one first grade classroom. During those 30-minutes around the reading table, children in the top group spend three times as much time on task as children in the bottom group, and McDermott has tried to understand how this happens. First, the procedure for allocating turns to read is different in the two groups. In the top group, the number of pages in the story is allocated equally among the children, and each child reads his share in order around the table. In the bottom group, there's no fixed order and each turn is negotiated according to who requests a turn and who the teacher thinks can read the page in question. Interruptions are more frequent in the bottom group (40 vs. 2 for the top group) and more disruptive because continuation of reading is more dependent on the teacher for assigning the next turn. Some of these interruptions are even initiated by the teacher herself:

> On one occasion, for example, she organizes the children to call for a turn to read their new books, "Raise your hands if you can read page 4." The children straighten themselves up in their chairs, form neat lines along the sides of the reading table, and either raise their hands for a turn or at least look at their books or the teacher. As their hands reach their highest point, the teacher looks away from the reading group to the back of the room. She yells at one child in the top group, and then another child in the top group. The three children in the bottom group who raised their hands, lower them to the table. Another little boy who didn't have his hand raised thrusts his chair back away from the reading table and the teacher and balances it on its two back legs. The other two children in the group simply look down at their books. The teacher returns and says, "Nobody can read page 4? Why not?" Eventually the children recover, and someone gets a turn. But it all takes time.
>
> (McDermott, 1978)

How does this contrast come about? Possibly the teacher has been told somewhere that calling on children in a random order helps keep the attention of potentially more disorderly children. More importantly, McDermott (1978) suggests:

> What is driving this whole system? I don't think it is the negative expectations of the teacher. Rather, the children in the bottom group represent pedagogical and interactional problems for the teacher. Pedagogically, there is no doubt that it is easier for the teacher to practice reading with the children in the top group than to struggle with the process of teaching decoding to the children in the bottom group. And interactionally, there is the pressure of the competition between the groups and the scarred identities of the children in the bottom group. Even within the bottom group we hear claims of one child against another. ("Oh, you can't read." "Better than you.") Or we can point to a child in the bottom group who constantly calls for turns to read while, at the same time, appears to struggle to make sure that she does not get eye contact with the teacher.
>
> In response to all these problems, the teacher and the children in the bottom group make adaptations. In response to all these pressures they struggle to solve the pedagogical and interactional problems of coming to school not knowing how to read, of having a teacher who expects them to know how to read, of having a teacher who doesn't know how to overcome that they do not know how to read while she has twenty other children walking around the room, and of overcoming the pressure of having the other children taunt them for their performances. In response to all this, they make very specific adaptations. One adaptation is to make sure that no one child is isolated to read something too difficult. So the teacher uses the two different turn taking systems with the different groups, and this adaptation has the consequences already explicated.

McDermott (1978) concludes:

> Success in learning is best predicted by the time a child spends on a task; some may learn faster than others, but with time, almost any child can learn what has to be learned in school, if there are the proper organizational constraints for getting the child on task for the necessary amount of time. The question of why some children achieve more than others has been transformed into a question about the environments in terms of which some children get consistently organized to attend to school tasks in classrooms while others do not. . . .
>
> Certain children, who, for whatever reasons come to school behind their peers in the development of classroom skills, constitute both pedagogical and interactional problems for most teachers. Most teachers say of them that they are harder to teach; part of that reaction is that they need more of the teacher's time if they are to catch up with

> their peers. In addition, they must learn under the pressure of knowing that they are behind, generally in a classroom which allocates status in part on the basis of the children's intellectual ranking in the classroom. . . .
>
> Thus, the small differences between children in the early years of school expand quickly to the drastic forms of differential performance which become obvious in later years. At the root of these differences is not so much the extreme complexity of the school tasks, nor the differences in the learning potentials of the different children, but the differential environments we offer the children for getting organized and on task so that learning can take place.

I think we have to acknowledge that what McDermott has exposed would be found elsewhere if we dared to take as close a look.

Fortunately, we have reports of one success story too. The Kamehameha Early Education Program (KEEP) is in a privately supported school for ethnically Hawaiian children, whose reading achievement in regular classrooms traditionally has been very low. In 1976 a new reading program was introduced at KEEP, and reading scores of the first grade children increased from an average of the 19th percentile in the preceding three years to the 69th percentile (Jordan, Weisner, Tharp, & Au, 1978).

According to Au (in press), the new reading lessons have three component parts, which she labels ETR, for *E*xperience, *T*ext and *R*elationships: The teacher begins by evoking comments from the children about their *experiences* that relate to the story (which is usually from a basal reader); she then assigns a page or two of *text* to be read silently and questions the children about the text; finally, she draws out *relationships* between the text and their experiences.

So far, except for the careful attention to evoking the children's personal experiences to engage their attention and provide schemata for comprehension, this sequence does not sound different enough to account for such striking gains. Au believes that the success lies not only in the cultural congruence of the content but, as with Piestrup's Black Artful teachers, in the cultural congruence of the context as well. Briefly, the rapid interactions between teacher and children, and the cooperative interaction among the children who build on one another's responses, produce lesson talk with striking similarities to "talk-story," a form of collaborative narrative of personal experience that is a special speech event in Hawaiian culture. Au (in preparation) is now documenting in more detail, from an analysis of videotaped lessons, how this talk happens and how the teacher channels the talk-story-like ways of speaking toward academic ends. Quantitatively, the children are certainly more engaged in reading tasks in the new program; but qualitatively the focus of their attention has been changed as well.

Hess and Takanishi (1974) observed student "engagement" in eight 30-minute observations in 39 elementary school classrooms in low-income communities to find out what teachers did to "turn on" their students to academic work in mathematics and language arts. Overall, they found that student engagement was strongly and consistently related to teacher behavior, but not to classroom architecture, nor to student characteristics such as sex and ethnicity. Two demonstrations of intra-teacher consistency in their data are impressive. First, two teachers were observed during two consecutive years. Although they had completely different classes and reported that they felt large differences between the two years, the mean level of engagement in their classes remained almost identical. Second, during the second year of the study, an entire school being observed moved from a self-contained classroom building to one with open-space architecture. The overall level of engagement across these very different physical environments was identical (82 and 83 percent), and the rank order of teachers in terms of percent engagement in their classrooms was .85.

Contrary to expectations, Hess and Takanishi found that these levels of student engagement were not consistently related across teachers to "specific teacher strategies" such as the frequency of specific questions or of feedback; instead they were strongly related to more "global instructional strategies" such as instructional group size (more engagement in small groups), and direction of student attention (more engagement when directed toward the teacher than toward other students or materials alone). The authors conclude with a recommendation that teacher-training programs concentrate on skills in classroom social organization rather than on more specific teaching behaviors. This is an important caution for competency-based training as it is usually conducted.

An observational study, done at the request of Children's Television Workshop, also measured children's engagement—or attention as we called it (Cazden, 1973). We wanted to find out what environmental variables affected the viewing behavior of children watching *The Electric Company* in their elementary school classrooms. Viewing behavior was defined as both visual attention and verbalizations. We observed ten primary grade classrooms during the 30-minute show five or six times each. Two independent measures of attention were used: a scan of the entire class at 30-second intervals to count those visually oriented toward the TV screens, and continuous monitoring and recording of the visual attention of individual students on an event-recorder. Monitoring individual attention on the event-recorder was extremely reliable (.94 interobserver agreement), and group attention averages from the 30-second scans had high validity (average within classroom correlation of .94 between measures of group and individual attention). Coding verbalizations was more difficult (interobserver reliability attained only .84). The ten classrooms were selected to represent a range in classroom "structure," defined here as a continuum from classrooms where attention to the show was expected and enforced by the teacher ("high" structure) to classrooms where a range of competing activities was available ("low" structure). As expected, we found that classroom structure was positively related to both group attention (correlation .87) and individual attention (correlation .95). High structure affected all children, increasing their attentiveness and responsiveness to *The Electric Company*, so that poor readers in high structure classrooms had higher attention scores than better readers in low structure classrooms.

With the exception of one classroom, structure also correlated highly with average number of reading responses (correlation of .90 for nine classes, but only .38 for all ten). In the one exception, the most highly structured classroom was highest in amount of attention paid by the students but lowest in average number of reading responses. Since there was nothing in the level of reading ability in the classroom that would explain this anomaly, we think that some aspect of this teacher's classroom control (which we could not understand from our limited observations) discouraged overt reading responses.

Because *The Electric Company* is designed especially for children reading below grade level, we were also interested in the relationship between viewing behavior and reading level. Children's reading ability can be categorized according to their relative standing in their class (high, middle or low reading group) or ranked more absolutely according to standardized test scores. Average attention of children in the lowest quartile of achievement test scores was 79.1 percentile. While this was not as high as the 86.5 percentile and 90.2 percentile attention of the two middle quartiles (25-75 percentiles), it was higher than the 65.8 percentile attention of the best readers and was encouraging evidence that the show was reaching its intended audience. More interesting and surprising was a finding that, without exception, children of the same tested reading level showed less attention and more fluctuations in their attention (more distractions) when they are among the lowest readers in their classroom than when they are in relatively higher reading groups. The data are shown below for the six second grade classrooms for which fall standardized test scores were available.

Attention and Fluctuation of Children in High and Low Reading Groups

	Comprehension Quartiles[a]			
Class Standing	1-25	26-50	51-76	76-100
		Percent attention		
High	0. (0)	89.5(10)	90.9(9)	73.2(11)
Low	79.1(20)	79.0(4)	87.3(2)	49.4(5)
		Number of fluctuations		
High	0. (0)	51.2(10)	30.6(9)	44.4(11)
Low	50.6(20)	58.6(4)	57.5(2)	64.6(5)

Note. The data are from Cazden, 1973.

[a] Numbers in parentheses indicate the number of children in each cell.

Because our sample was not designed for matching numbers of children in each of these cells, firm conclusions cannot be drawn. But in these admittedly limited data, lower relative standing in class (in terms of reading group assignments) adversely affects children's attention to televised reading material. Seen in this way, a variable such as reading level (usually considered a child variable in its absolute sense) becomes an environmental variable as well through the child's relative standing in the classroom group.

Differences in the Focus of Instruction

Time on task is a powerful variable, but it is not the only one. One more qualitative variable is where the attention of children and teacher is focused during reading instruction. We know that learning to read requires mastering a complex set of concepts and skills at many levels of a hierarchical system. Analytically, we can separate a series of nested units—from the meaningless sounds symbolized by letters, to larger and larger meaningful units of words, phrases, clauses, paragraphs and stories; and we can isolate the conventions of punctuation, capitalization and layout on a page that support the communication of meaning (remember that the division of print into lines is one visual feature that does not carry meaning except in poetry and that children must therefore learn to ignore). But such analytical separation says nothing about how children should be helped; it does not determine in what order their attention should be focused on different units in the hierarchy, nor how an eventual integration can best be achieved.

The simplest contrast in focus is between decoding skills and meaning. We know we cannot tell what actually happens from the manuals on a teacher's desk or the methods she professes to use. For example, in one of the first grade reading studies supported by the Office of Education Cooperative Research Project, Chall and Feldman (1966) went behind "method A vs. method B" comparisons to examine what teachers actually did to implement those methods. Observational studies of teachers showed no significant relationship between the ranking of the teacher's professed method emphasis (whether "sound-symbol" or "meaning") and the method emphasis observed in her classroom (p. 573).

If attention to phonic skills and to meaning is included in reading group lessons, then that combination can create problems of shifting focus and complex interactions that will be discussed further below. Here I want to report research that describes classrooms in which these foci are separated— by children, by type of instructional event, or by language.

Separation of Focus by Children

In the classroom studied by McDermott, the focus differed from one group in the classroom to another.

> [In the top group] occasionally, the children create problems by word calling instead of reading for meaning, and the teacher's main pedagogical task is to convince the children that there is living language complete with propositions with illocutionary force on the page. Thus, one child reads, "But Ricky said his mother. . ." in a dull monotone, and the teacher corrects her, "Let's read it this way, 'But Ricky, said his mother'."
>
> With the bottom group, the teacher has rather different problems. Accordingly, the teacher and the children constitute rather different environments for each other in the different groups. The children in the bottom group do not read as well as the children in the top group, and the teacher attends less to the language on the book's pages and more to the phonics skills needed to interpret any given word in the text. Thus, there are many more stopping places in the children's reading, and the story line which is to hold the lesson together is seldom alluded to and never developed. (McDermott, 1978)

This same contrast between focus on meaning for better readers and focus on phonic skills for poorer readers is found in two other studies by Gumperz (1972) and Allington (1978). Gumperz reports observations in a first grade classroom in a racially integrated California district:

> We observed a reading session with a slow reading group of three children, and seven fast readers.... With the slow readers she [the teacher] concentrated on the alphabet, on the spelling of individual

> words.... She addressed the children in what white listeners would identify as pedagogical style. Her enunciation was deliberate and slow. *Each word was clearly articulated with even stress and pitch....* Pronunciation errors were corrected whenever they occurred, even if the reading task had to be interrupted. The children seemed distracted and inattentive....
>
> With the advanced group on the other hand reading became much more of a group activity and the atmosphere was more relaxed. Words were treated in context, as a part of a story.... There was no correction of pronunciation, although some deviant forms were also heard. The children actually enjoyed competing with each other in reading and the teacher responded by dropping her pedagogical monotone in favor of more animated natural speech. (Gumperz, 1972)

Allington's study (1978) suggests that this contrasting focus is not just a chance characteristic of two classrooms that happened to be studied by McDermott and Gumperz. Allington audiotaped oral reading segments of reading lessons with the best and the poorest readers in 20 primary classrooms in three school districts. He analyzed how the teachers responded to children's oral reading errors and found dramatic differences between the two reading groups across the 20 classrooms, differences which fit exactly the pictures described more qualitatively by McDermott and Gumperz. First, there was a difference in the rate of teacher corrections of the errors (68 percent of poor readers' errors were corrected, but only 24 percent of good readers' errors). Second, there were differences in the timing of the correction: teachers were more likely to interrupt poor readers at the point of error (88 percent of poor readers' errors vs. only 70 percent of good readers' errors) rather than waiting for the next phrase or clause boundary. Finally, there were differences in the cues provided by the teachers to help the children read the right word: for the poor readers, the cues were more apt to be graphemic/phonemic (26 percent vs. only 17 percent of the cues for good readers), while the cues for good readers were more apt to be semantic/syntactic (31 percent vs. 14 percent for poor readers).

The critical question raised by these reports is whether such differentiated teacher behavior is helpful or not. Marie Clay speaks from New Zealand of the goals of education as helping children become "self-improving systems" (personal communication). In other words, the goal is not to create children who never make mistakes, but rather children who have the capacity to notice their own mistakes and have strategies for correcting them. She has found that children in the first grade who do the most self-correcting are the children who become the better readers in late grades (Clay, 1973). Does being interrupted make self-correction more or less likely to develop? And if a cue from the teacher is needed, what kind of a cue should it be?

Allington's paper is titled, "Are good and poor readers taught differently? Is that why poor readers are poor readers?" The implication is clear that he believes it is possible (as do I) that these teacher behaviors to low group children may increase their problems in the long run. Prompt interruptions seem too much like a "law and order" approach to errors, as if the teacher is acting out of fear that the errors, like the children themselves, may get out of control. But just because it is the long run that counts, we need longitudinal studies that follow teacher behaviors and children's progress over time. (I am grateful to Rebecca Barr for this caution.) Only then can we separate constructive individualization from destructive bias.

Separation of Focus by Instructional Event

A very different kind of separation of foci of attention is by instructional events distributed throughout the school day. As part of a larger study of children's functional language competence in kindergarten and the primary grades conducted at the Center for Applied Linguistics, Griffin (1977) has described the set of events in which reading happens in one first grade classroom. These include: reading a recipe for hot cross buns that leads to a discussion of the meaning of "lukewarm" and experiments with feeling lukewarm milk later in the day; and story time when the teacher reads aloud, stopping frequently for talk about what is going to happen next. Griffin notices that comprehension skills of vocabulary and prediction were built in such nonreading group times of the day; whereas in the reading groups themselves, phonics was the overriding concern. This separation was so consistent that definite expectations about appropriate responses had been learned by the children. If the teacher shifted momentarily to a meaning cue during a reading group, the children were apt to respond with a phonic-based response anyway. For example:

> One child was reading, in a very halting style: "The pigeon flies far...." He paused. The teacher repeated the sentence in a more fluent style with correct intonation and then gave the child a prompt: "The pigeon flies far.... Think. Think what it would say! The pigeon flies far...." A second child chimes in saying, " 'A' says it name. *Away*." (Griffin, 1977, p. 381)

In considering the merits of such a separation by instructional events, we must remember that it can only work in classrooms where there is a rich set of nonreading group events in which reading takes place. One tragic result of the pressures of the back-to-basic movements may be less time available for experiences like reading recipes and hearing stories in which vocabulary building and comprehension education can so meaningfully occur.

Separation of Focus by Language

In the most extreme case, a focus on meaning and a focus on phonics may be separated in a single classroom not only by reading events, but by languages as well. While this review does not attempt to cover research on learning to read in two languages, my own observations in a bilingual first grade classroom in Chicago are relevant here.* In the fall, the teacher's reading instruction was in Spanish only, using a traditional syllabic approach (ma me mi mo mu). Around Christmas, when she felt that the children's oral English had developed sufficiently, the teacher started a phonic-based reading program in English. As she described that attempt afterward, it just didn't work; the children resisted and she finally stopped. About that time, in a graduate class she was taking, she read Sylvia Ashton-

---

*These observations are part of a research project on "The social and cultural organization of interaction in classrooms of bilingual children," supported by NIE grant 780099 to Frederick Erickson and Courtney B. Cazden.

Warner's *Teacher* and felt immediately that those ideas fit her philosophy and her children. The result was that by February, when I visited again, instruction in Spanish reading via syllables coexisted, for all children, with instruction in English reading via "key words." Moreover, the teacher was consistent in the cues she gave the children in the two contexts. In Spanish she focused their attention on the syllabic components on which they had had extensive practice:

Fe li pe to ma u na fo to.

In English, she helped with a meaning cue: when a child couldn't read *butter* on his key word card, she asked, "What do you put on your toast?"

At first thought, such separation may seem detrimental to learning. Intuitively, it seems harder for children to get decoding and meaning cues together in a single mental act if they are taught separately in different parts of the school day, or even in different languages. On the other hand, maybe a clear and consistent focus of attention is helpful, especially for beginning readers, as long as both are included somehow, for all the children.

The Complexities of Reading Group Lessons

Primary grade reading groups are complex interactional scenes—complex because of the triangular relationship between a reader, a text being read, and the participation of teacher and peers; because of the many levels of organization of the text that may move unpredictably in and out of the focus of the teacher's instruction; and because oral reading serves simultaneously as practice for the child and a context for evaluation by the teacher. Two research reports, by Dickinson, Kozak, Nelson, and Epstein (1977) and Heap (1978, 1979) say more about these complexities.

Dickinson et al. described differences in single vs. multiple foci, and attendant differences in time spent off-task, in a math lesson and a reading group lesson with first grade children in a single K-1 classroom. In the math lesson, the children were individually manipulating attribute blocks into intersecting sets. There was a repeated and, therefore, predictable sequence of teacher directives about placement of the blocks, questions to the children about what they had done, and finally a concluding statement about what they had found out. In successive sequences, the two parts of each directive (e.g., "Place the blue blocks in this circle" and "Place the yellow blocks in this circle") were spoken with decreasing intervening time, and successive questions to the children elicited progressively more information. In the reading group, in contrast, there was more variation and less predictability in both the focus of attention and the interactional structure. The teacher asked individual children to take turns reading aloud, but talk about the book title, table of contents, page numbers, and capital vs. lower case letters was interspersed in seemingly unpatterned ways.

There were so many other differences between the two groups that no firm conclusions can be drawn—differences in activity, group size, and whether the group included all children present or only a subset. While the reading group was smaller, it did not include all the children in the room at the time and so was more subject to interruptions and divided teacher attention. It would take more controlled research to determine how much the interactional simplification of the math group alone contributed to the greater on-task engagement.

The possible instructional value of such interactional simplification is not a new idea. Some of the success of Distar may be due to this feature. Such simplification has also been advocated in a discussion of the design of Sesame Street (Gibbon, Palmer, & Fowles, 1975). A familiar example of holding the instructional frame constant while varying the content is the Sesame Street categorization game, "One of these things is not like the other." Gibbon et al. explain the reason for this design:

> Varying the content while keeping the format constant promotes familiarity with format conventions that are potentially useful for instructional purposes. The format of any program segment functions as a kind of "frame" for the instructional content, a complex of auditory and visual conventions that the child can master through repeated exposure. For example, the viewer can learn to expect that a particular format will usually deal with a particular category of stimulus (letter, word number, concept) and with a particular intellectual activity (memorizing, sorting or classifying, guessing, combining). A particular sequence of events or types of events will reliably occur; a particular

> type of feedback to the viewer's implicit or explicit responses will be delivered. Moreover, a viewer's familiarity with a given format can help him determine at what point in the presentation the important information will come, how much of it there will be, perhaps even whether it is likely to be too easy, too difficult, or about right for him. Among the main instructional advantages afforded by these various forms of cueing is that they will entice the viewing child to attend to what is new in each succeeding application of the format, since it will "stand out" against the familiar background more than if the entire presentation were novel. As a result, learning and concept formation are enhanced. (Gibbon, Palmer, & Fowles, 1975, pp. 225-226)

Reading groups as traditionally enacted in primary school classrooms are inherently complex in content and interactional structure because learning to read requires so many different kinds of learnings. We need interactional analyses of alternative organizations of reading events in which these learnings are separated or combined.

Heap (1978, 1979) is studying the "social organization of reading activities" in 20 classrooms. He has finished only one year of a five-year project, and so only preliminary reports are available. In these reports, he has identified three "social organizational problems" in primary grade reading instruction; two will be familiar to teachers, and all three raise important questions about the relationship between social organization and individual cognition.

The two familiar problems are problems in evaluating a child's response. On the one hand, a child's correct answer in a reading group lesson may be an artefact of other resources that the group provides. As Heap describes a specific example, "As a task organized to make reading skills observable and evaluable, the reading lesson provided an unforeseen resource, reading aloud, for a participant to continue to participate in the task" (Heap, 1979, p. 4) by answering comprehension questions even though her book had been closed. On the other hand, a child's reading errors may be due to obstacles created by that same reading group organization—for example, anxiety about the social performance of reading aloud in front of peers.

The third organizational problem is more complex. Here is Heap's example, from the comprehension section of a second grade reading group lesson after the first part of "Rumpelstiltskin" had been read:

Teacher	No. Who helped Mineen?
Child	Rumpelstiltskin.
Teacher	Yeah the little man. We don't know his name is Rumpelstiltskin yet do we? The little man. Okay, what was the first thing the prince said—sorry, that the girl gave to Rumpelstiltskin, to the little man. We better call him the little man because we don't know really he's Rumpelstiltskin yet (Heap, 1978, p. 2).

The story was titled "Rumpelstiltskin"; the teacher had written that name on the board as a new vocabulary word at the beginning of the lesson; and she knew that several of the children had seen a movie version of the story the previous year. Yet she still corrected the child, and self-corrected herself, from saying "Rumpelstiltskin" to the vaguer "little man." As Heap says, it is only true that "we don't know his name is Rumpelstiltskin yet" in a very special sense: within the limits of and the terms of a convention, a game, that disengages reading and answering questions about that reading from everything else the child knows, from everything outside the boundary of the text itself.

These "organizational problems" that Heap has described are not unique to reading groups. Any exercise of any cognitive process—for us, as for children—takes place in some context: of particular task format, physical conditions, social organization, conventional rules, etc.; and characteristics of that context will contribute either supports or obstacles to the cognitive tasks performed within it. Cognition is always in some context; and it can't be taught or evaluated apart from some particular context either. (I am indebted to Michael Cole for many conversations on this point.) Because of the importance of reading groups as a context for both instruction and evaluation in the primary grades, we need to examine that context—and the "games" we play within it—with particular care. Yet because reading groups are so traditional and so familiar, that examination is especially hard for participants to do for themselves. To complete the circle back to the first study by Hess and Takanishi reported at the beginning of this review, classroom teaching must be considered as a complex orchestration of social life in which diverse individual cognitive processes can most effectively be developed, and we need to understand teaching from both the social and the cognitive points of view.

Peer Group Interaction in the Older Grades

This review has focused on learning to read in the primary grades—partly because of my own interests and experiences as a teacher (Cazden, 1976) and researcher with younger children, partly because most of the recent sociolinguistic and ethnographic research has focused on the primary grades as the place where children are first inducted into the school "culture" and where their academic "identities" are first formed. I assume that classroom interactions are not less important for reading instructions in the intermediate and secondary grades, but that the relevance takes different forms. One important form is the relationship between the instructional process and interactions among peers.

Consider the implications of just one study: Labov's research on the relationships among the incidence of nonstandard Black English (BE) dialect features, peer group membership, and reading failure (Labov & Robins, 1969; Labov, in press). Labov and his colleagues analyzed the incidence of BE features in the speech of black adolescents in fourth through tenth grade, identified the speakers as either central or peripheral members of peer group in the street culture by sociometric interviews and participant observation, and then correlated these data with performance on standardized reading tests.

One linguistic indicator of BE dialect is saying *have* for *has* in the third singular present. Labov (in press) reports the frequency of the standard form *has*, and then comments on the group differences:

> ...club members used only 19 percent of the *has* form; the lames [isolates from the street culture] used 60 percent; and white working-class adolescents 100 percent.... These [dialect] differences are slight: they are small differences in the probability of a rule being applied. They reflect patterns of communication and ideology, but in no way could they be conceived of as the causes of differences in reading achievement.

Differences in reading achievement existed. All 46 club members, core participants in the street culture, reached a virtual plateau in reading achievement at the fifth grade level, while the 32 lames continued to progress, one-third of them at or above grade level. According to Labov, the dialect differences are not in themselves the cause of these reading problems; they are rather the indicators of group membership and of a value system that accompanies such membership that is in conflict with the school.

Support for Labov's argument comes from the co-occurrence of events around the fifth grade watershed year. This is the time when peer group formation differentiates members and lames in life on the street outside of school, and also differentiates their reading achievement within. More generally, fifth grade is the point at which, across the country, poor children's reading scores decline relative to their richer peers:

> A new state report [by the state legislative analyst's office] says the achievement gap between richer and poorer children in the California schools appears to be widening.... [this decline] seems to be part of a national trend, with the decline beginning around the fifth grade and increasing later. (*Palo Alto Times*, 11/3/78)

While these reading achievement data are usually interpreted in terms of the changing character of reading texts and tasks in the intermediate grades, it seems probable that, for some students, value conflicts accentuate the problem.

The connection between this research and the focus of this review comes in Labov's (in press) discussion of possible remedies.

> The techniques of learning and studying imposed by our schools are avowedly individualistic and competitive. Each student is expected to learn by himself, and as I noted at the outset, interaction in the classroom is fundamentally confined to dealing directly with the teacher....
>
> The skills that are highly developed in vernacular culture depend upon a different strategy. Sports, formal and informal, depend upon close cooperation of groups. The same holds for music.... Individuals practice by themselves, but the major steps in learning are done in tempo with the group....
>
> If we continue to repress vernacular culture, and try to extract one or two individuals from their cultural context, we will continue the pattern

> of massive educational failure that we now observe in the schools. The other route is to understand the interests and concerns of the youth who come to school and use that understanding in a positive way (Labov, in press)

This "positive way" will have to make possible less individualistic ways of learning to read, so that the power of group interactions can be used directly as contexts for learning. (See Steinberg & Cazden, in press, for one report of peer teaching in an intermediate grade.) Unfortunately, in discussions of teaching, the term "classroom interaction" has become limited to interactions in which the teacher is involved. We need to expand its meaning back to include all interactions in which learning takes place—not only with the teacher, but among students as well.

References

ALLINGTON, E.L. *Are good and poor readers taught differently? Is that why poor readers are poor readers?* Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

AU, K.H. ETR: Start with the experiences of the minority culture child. *The Reading Teacher*, in press.

AU, K.H. *A test of the social organizational hypothesis: Relationships between participation structures and learning to read.* Unpublished doctoral dissertation, University of Illinois, in preparation.

CAZDEN, C.B. *Watching children watch 'The Electric Company': An observational study in ten classrooms.* Final report to Children's Television Workshop, August 1973.

CAZDEN, C.B. How knowledge about language helps the classroom teacher—or does it: A personal account. *The Urban Review*, 1976, *9*, 74-90.

CHALL, J., & FELDMAN, S. *A study in depth of first grade reading: Analysis of the interactions of proposed methods, teacher implementation and child background.* Cooperative Research Project No. 2728, U.S. Office of Education, 1966.

CLAY, M.M. *Reading: The patterning of complex behavior.* Auckland, N.Z.: Heinemann, 1973.

DICKINSON, D., KOZAK, N., NELSON, E., & EPSTEIN, M. *Examination of differences in the dynamics of small and large group instruction.*

Unpublished term paper, Harvard Graduate School of Education, 1977.

GIBBON, S.Y., JR., PALMER, E.L., & FOWLES, B.R. Sesame Street, The Electric Company and reading. In J.B. Carroll & J.S. Chall (Eds.), *Toward a literate society*. New York: McGraw-Hill, 1975.

GRIFFIN, P. How and when does reading occur in the classroom? *Theory into Practice*, 1977, *16*, 376-383.

GUMPERZ, J.J. Verbal strategies in multilingual communication. In *Georgetown University Round Table on Languages and Linguistics 1970*. Washington, D.C.: Georgetown University Press, 1972.

HEAP, J.L. *Rumpelstiltskin: The organization of preference in a reading lesson*. Paper presented at the annual meeting of the Canadian Sociology and Anthropology Association, London, Ont., June 1978.

HEAP, J.L. *The social organization of reading evaluation: Reasons for eclecticism*. Toronto: Ontario Institute for Studies in Education, April 1979.

HESS, R.D., & TAKANISHI, R. *The relationship of teacher behavior and school characteristics to student engagement* (Tech. Rep. No. 42). Stanford, Calif.: Stanford Center for Research and Development in Teaching, 1974.

JORDAN, C., WEISNER, T., THARP, R.G., & AU, K.H. *A multidisciplinary approach to research in education: The Kamehameha Early Education Program* (Tech. Rep. No. 81). Honolulu: The Kamehameha Schools, 1978.

LABOV, W., & ROBINS, C. A note on the relation of reading failure to peer-group status in urban ghettos. *The Teachers College Record*, 1969, *70*, 395-405.

LABOV, W. Competing value systems in the inner-city schools. In P. Gilmore & A. Glatthorn (Eds.), *Ethnography and education: Children in and out of school*. Philadelphia: University of Pennsylvania Press, in press.

MCDERMOTT, R.P. Pirandello in the classroom: On the possibility of equal educational opportunity in American culture. In M.C. Reynolds (Ed.), *Futures of exceptional children: Emerging structures*. Reston, Va.: Council for Exceptional Children, 1978.

PIESTRUP, A.M. *Black dialect interference and accommodation of reading instruction in first grade* (Monographs of the Language-Behavior Research Laboratory No. 4). Berkeley: University of California, 1973.

STEINBERG, Z., & CAZDEN, C.B. Children as teachers—of peers and ourselves. *Theory into Practice*, in press.

Social-Psychological Perceptions and Reading Comprehension

Herbert J. Walberg
Victoria Chou Hare
Cynthia A. Pulliam
University of Illinois at Chicago Circle

Ancient psychologists and philosophers recognized the need for a balanced understanding of human beings in terms of the head, the heart, and the hand, or in modern psychological terminology—cognition, affect, and volition; but it has often been said that psychology lost its soul a long time ago, destroyed its mind at the turn of the present century, and is now in trouble with behavior. Moreover, although psychologists have made valiant efforts to apply laboratory findings to natural settings of human action, the results have often been discouraging; for example, the so-called behavioral and cognitive revolutions in psychology are not notable for raising achievement test scores in ordinary classrooms in recent decades. What has been lacking in much psychological theory and research on school learning is a consideration of the student perceptions of the social-psychological environment of their classes, and the direct and indirect linkages of these perceptions to student outcome measures—such as standardized test performance, interest in the subject, and self-concept as a learner—that are of interest to public policy makers, educational practitioners, parents, and students.

About a decade ago, however, research on student perceptions of the social-psychological aspects of their classes began in high school physics in the United States and then expanded to other subject matters, grade levels, and countries. The research shows that student perceptions can be validly measured and can serve as an index for classes or individual students of the extent of cognitive, affective, and behavioral learning that goes on during the school year or shorter periods of time. The purposes of this chapter are: 1) to explain how these student perceptions can be conveniently measured in ordinary classrooms, 2) to review the research on the prediction of learning outcomes from student perceptions, 3) to relate this research to models of reading comprehension, and 4) to discuss the implications of the first three sections for research on reading as well as the practice of teaching reading.

Research on student perceptions comes under the headings of morale, climate, or social environment in recent educational research. These terms emphasize that it is the general perception of the composite qualities of classes rather than their so-called objective characteristics such as group size and counts of collective behavior that are being examined. There are a great number of theoretical and methodological issues that cannot be fully discussed here because of space limitations, but the interested reader may find the following publications of help in pursuing the topics briefly considered here: Insel and Moos (1974) on social-psychological research of perceptions of a variety of human settings; Campbell (1970) and Majoribanks (1974) for valuable but neglected collections of substantive research on learning environments carried out in Australia, Canada, England, and the United States; Kahn and Weiss (1973), Randhawa and Fu (1973), Shulman and Tamir (1973), and Walberg (1974, 1976) for substantive and methodological reviews; and Walberg (1974, 1979) for sourcebooks on a variety of learning environment instruments, studies, and evaluations by a number of international groups.

Measurement and Validity of Student Perceptions

Much of the research on student perceptions of their classes employed the Learning Environment Inventory (LEI) or the My Class Inventory (MCI). Table 1 shows sample items for the

Table 1
Sample Items, Reliabilities, and Consistencies of 15 Variables

	Sample Item	Class Reliability	Percent Positive Correlations with Learning
Affect Variables			
Cohesiveness	All students know each other very well.	.84	86
Satisfaction	The students enjoy their class work.	.79	100
-Friction	Certain students are responsible for petty quarrels.	.80	0
-Apathy	Students don't care what the class does.	.77	14
-Cliqueness	Some students refuse to mix with the rest of the class.	.74	8
Status Variables			
Democracy	Students have about equal influence on the class.	.61	85
Competitiveness	There is much competition in the class.	.56	67
-Diversity	Interests vary greatly within the group.	.36	30
Favoritism	Certain students are favored more than the rest.	.65	10
Task Variables			
Goal Direction	Each student knows the goals of the course.	.73	73
Material Environment	The room is bright and comfortable.	.79	86
Formality	The class has rules to guide its activities.	.87	65
Difficulty	Students find the work hard to do.	.81	87
Speed	The course material is covered quickly.	.76	54
-Disorganization	The class is disorganized and inefficient.	.87	6

Note. The learning criteria include cognitive, affective, and behavioral measures including higher mental processes and self-concept.

LEI intended for students in junior and senior high schools. The LEI consists of 105 items, 7 items per scale on each of 15 scales. Students rate their classes on four points—strongly agree, agree, disagree, and strongly disagree—on each of the items. For example, on the Cohesiveness scale, a student is asked to agree or disagree with the item "All students know each other very well." The seven items on each scale are summed for analysis.

The LEI can be administered in approximately 25 minutes and it produces reliable estimates of the morale, climate, or learning environment of the class. In the interest of saving student time, some investigators have omitted some of the scales or reduced the number of items per scale from 7 to 3 without too much sacrifice in reliability. The scales have been translated into several modern languages and used in a number of countries in Africa, the Americas, Asia, and Europe.

The My Class Inventory is an adaptation of the LEI for elementary school research. It consists of 45 items, 9 items per scale on each of 5 scales. The vocabulary level of the items has been reduced; but it is sometimes found to be necessary to read and explain the items to some students and explain items to classes of poor readers. Working from these two standardized instruments, a number of investigators have modified the items and scales to suit them to their particular research interest and setting.

Prediction of Learning Outcomes

From social-psychological research, Walberg (1969) formulated 36 hypotheses concerning the direction of relations between selected LEI scales and learning criteria, namely that Cohesiveness, Satisfaction, Task Difficulty, Goal Direction, Democracy, Diversity, and Material Environment would be positively correlated, and that Friction, Cliqueness, Apathy, Favoritism, and Disorganization would be negatively correlated with the extent of gains in cognitive, affective, and behavioral learning. A tabulation of the results (Haertel, Walberg, & Haertel, 1979) across 10 large data sets shows that 31 out of 36 or 86 percent of the signs support the formulations (Table 1); the probability of these results occurring by chance is less than .001. Three of the five disconfirmations concern the Diversity scale which shows generally negative relations with outcomes in all three learning domains, rather than the hypothesized positive direction. The results which are tabulated in the last column of Table 1 show the general consistency of the findings in the research done in the past decade.

To estimate the sizes of the correlations of student perceptions with learning outcomes and adjusted learning gains, 734 correlations from a comprehensive collection of 12 studies of 10 data sets on 823 classes containing 17,805 students in eight subject areas and four nations were synthesized (Haertel, Walberg, & Haertel, 1979). Table 2 shows a summary of the findings for two specific areas—high school history and elementary school reading. Most of the investigations upon

Table 2
Correlations of Class-Mean Student Perceptions with Gains in Learning

	High School History		Elementary School Reading	
	Estimated	Observed	Estimated	Observed
Affect				
Cohesiveness	.38	.81	.17	.00
Satisfaction	.45	.63	.38	.11
-Friction	-.80	-.90	-.52	-.37
-Apathy	-.33	-.86		
-Cliqueness	-.27	-.74		
Status				
Democracy	.36	.61		
Competitiveness	.19	-.15	-.23	-.51
-Diversity	.19	-.15		
-Favoritism	-.31	-.59		
Task				
Goal Direction	.39	.66		
Material Environment	.55	.80		
Formality	.23	-.12		
Difficulty	.28	.26	-.14	-.27
Speed	.19	-.15		
-Disorganization	-.25	-.71		

Note. The estimates are based on a quantitative synthesis of more than 700 correlations from 10 independent data sets (Haertel, Walberg, & Haertel, 1979).

which the estimated correlations are based obtained pretest measures of achievement and interest in the subject during the first month of the fall term, obtained similar posttest measures at the last month of the spring term, and measured student perceptions of the social-psychological environment of their classes during the middle of the school year.

The correlations of student perceptions with adjusted gains in learning vary from low to high across variables and investigations; Table 2 shows examples from two studies. In the area of affective relations in the class, Friction in the group, that is, tensions and hostility among class members, is negatively correlated with learning outcomes; and Satisfaction and Cohesiveness are positively correlated with learning outcomes. In the general area of status within the classroom group, the extent to which the class is seen to be democratic and not favoring particular individuals in the group is moderately correlated with learning efficiency. In the area of task orientation, the perceptions of Goal Direction and the presence of materials and supplies in an adequate physical facility for carrying out the activities of the class are positively correlated with learning outcomes; and Disorganization is negatively correlated with learning gains.

When the scales are weighted and added to form a composite index of the social-psychological environment conducive to learning rather than using one scale at a time, the predictive validity increases substantially. Seven studies added sets of scales to regression equations that also included ability or pretest measures or both as controls. The average incremental variance accounted for in learning outcomes beyond that accounted for by ability and pretest on 19 learning outcomes is 20 percent, with a range from 1 to 54 percent; these incremental variances are substantially greater than that accounted for by IQ when pretests are taken into account. Thus, regressions containing the control and perceptual variables account for a large amount and, in some cases, nearly all of the reliable variance in learning outcomes. These analyses show that the measures of the social-psychological environment during the course of learning afford a very accurate prediction of how much will be learned during the school year and serve as a useful index of the average amount that the class is learning at any given time.

Perception and Reading

Although learning outcomes and gains are positively associated in general with Cohesiveness, Satisfaction, Task Difficulty, Formality, Goal Direction, Democracy, and Material Environment and negatively associated with Friction, Cliqueness, Apathy, and Disorganization, 8 of the 10 data sets have been gathered in various countries from secondary school classes in the sciences, humanities, and social studies. Only two studies have been carried out in the elementary grades, and only one of these has concerned reading.

Talmage and Walberg (1978) showed that student perceptions of their reading classes yielded the most accurate predictions of end-of-the-year standardized reading test scores with adjustments for similar pretests among four sets of

Table 3

Correlations of Learning Gains with Four Sets of Variables

Reading series	
Ginn 360	.00
Holt, Rinehart, and Winston	.08
Lippincott	.22*
Scott, Foresman Systems	-.23*
Teacher characteristics	
Years in teaching	.22*
Workshops attended	-.01
Reading courses	-.13
Awareness of design	.12
Instruction	
Decisionmaking	.14
Student behavior	.03
Materials utilization	.09
Environment	
Cohesiveness	.00
Competitiveness	-.51***
Difficulty	-.27**
Friction	-.37***
Satisfaction	.11

Note. One, two, and three asterisks mark correlations respectively significant at the .10, .05, and .01 levels.

prediction variables. Table 3 shows that, for a sample of 60 classes, grades one through six, the reading series used, teacher characteristics, and the nature of instruction were very weakly correlated with adjusted learning gains. Competitiveness, Difficulty, and Friction were moderately correlated in a negative direction with the extent of gains on standardized reading tests.

A selection of items on the My Class Inventory for these scales illustrates the characteristics of classes in which poor reading gains were made. On the Friction scale, several items are: "Some of the children in our class are mean." "Children are always fighting with each other." and "Some pupils don't like other pupils." On the Competitiveness scale are the following items: "The same people always do the best work in our class." "Children often race to see who can finish first." and "Most children want their work to be better than their friends' work." On the Difficulty scale are the following items: "In our class the work is hard to do." "Only the smart people can do the work in our class." and "Children often find their work hard."

These results are important for several reasons. First, they show that the general predictive validity of student perceptions found in secondary classes is initially confirmed in elementary school. Second, they call attention to the need for further research on social-psychological perceptions in elementary school reading. Although Friction is negatively associated with learning outcomes in elementary reading and thus confirms the results from secondary classes, the other two scales are reversed in their signs; that is, although Competitiveness and Difficulty are weakly and positively related to learning gains in secondary classes, they are negatively correlated in the sample of elementary school students with reading achievement as the outcome. These reversals in sign may be attributable to the idiosyncracies of a single study, or they may indicate that substantially different social-psychological environments are conducive to learning in elementary school. For example, it may be that difficult, highly competitive norms are conducive to learning in high school students but deter reading gains in elementary school students.

As mentioned earlier, the My Class Inventory contains 5 of the 15 scales from the Learning Environment Inventory; it is a matter of speculation how the other 10 scales will relate to learning outcomes in elementary schools. From the research synthesis on learning environments and from prior social-psychological research in laboratory and natural settings of learning, it seems reasonable to infer that subsequent research in all levels of education will continue to show that Cohesiveness, Satisfaction, Goal Direction, Democracy, and the Material Environment enhance learning, and that Friction, Cliqueness, Apathy, and Disorganization deter learning. However, additional studies in elementary schools and in the field of reading, in particular, are urgently needed to confirm or disconfirm these inferences.

Perceptions as Goals

Because learning environment scales provide a predictively valid index of the amount of learning gains made during the academic year as indexed by standard tests, the scales can occasionally serve as convenient substitutes for the standardized achievement and comprehension tests themselves. Since the scales are closely linked to affective and behavioral outcomes as well, they are useful indices in a variety of settings and subject matters. At the secondary level, the scales are fairly context free; that is, social-psychological perceptions conducive to learning in one set of circumstances are also conducive to learning in other sets of circumstances.

Because different curricula, teachers, and forms of instruction emphasize different sets of goals and outcomes of learning, standardized tests are not always an accurate or fair assessment across learning settings; and, in these circumstances, the LEI scales form a useful index of curriculum, instructional, and other kinds of educational effects. Since the earlier research demonstrated the predictive validity of the instruments, recent research has employed the scales to provide statistically-reliable sensitivity to educational treatments such as curriculum, teacher training, and instructional innovations, as well as to project efforts to increase student teamwork, cross-sex and cross-ethnic group cooperation, and similar properties. Other contemporary work reveals that student perceptions reflect and mediate teacher and student characteristics, and that they provide diagnostically-valuable profiles for class and individual morale or climate that can be used in needs assessments and impact evaluations. More than 600 investigators from many nations have requested the LEI and MCI scales for their work in evaluation and research.

In summary, scales measuring student perceptions of the social-psychological environment of learning provide useful independent, mediating, and dependent variables in educational investigations in natural settings. They complement and supplement current behavioral and cognitive variables and reveal social realities in classes that are neglected in laboratory-derived research. Even if the purpose of a research effort does not concern student perceptions themselves, the perceptions can be used as control variables so that subtle cognitive and behavioral effects may be detected in the complexities of natural settings. Lastly, it is the student that is the client of the educational system and the person that the system is attempting to influence. Surely these perceptions should be one basis of consideration and accountability in the evaluation of instruction, curriculum, and other aspects of the educational program.

This section illustrates the importance of student perceptions of their classes. The next section of this chapter outlines three models of reading which incorporate the dimension of comprehension. The last section links research on student perceptions to components of these models to hypothesize how comprehension is affected by the social-psychological environment in natural settings of learning.

Models of Reading

Many reading teachers and researchers would subscribe to the following description of reading comprehension by Gibson and Levin (1975): "We comprehend the meaning of a word, the meaning of a sentence, or the meaning of a passage of discourse when we apprehend the intention of the writer and succeed in relating his message to the larger context of our own system of knowledge" (p. 400). Like other descriptions of reading comprehension, this is very general and lends itself to different interpretations of how the process of reading comprehension actually operates. The following discussion briefly describes three much-quoted, yet very different, reading models and points out their similarities and differences in postulated structures and processes. Common and distinctive features of the three models provide leverage points for hypothesized influences of student perceptions of their learning environment on reading comprehension.

Current literature illustrates at least three types of reading models: bottom-up, top-down, and interactive. These designations refer to different assumptions about the nature of the reading process. If reading is initiated by visual stimuli or print, requiring little input from the reader, processing is referred to as bottom-up; Gough's (1972) model of the reading process exemplifies this type of model. If reading is initiated by reader-generated hypotheses, requiring fewer visual cues, processing is referred to as top-down; Goodman's (1970) model of the proficient reader adheres fairly well to the characterization. If reading requires bottom-up as well as top-down processing, the model is referred to as interactive; Rumelhart's (1975) work represents this type of model.

Despite their different characterizations of the reading process, certain structures and processes are common to the three models. All the models incorporate visual activities as initial processes, although only Goodman hypothesizes in his model that the reader must bring expectations about the nature of the reading task from both long- and medium-term memory structures to this process. The three authors differ in their notions about the kind of information which is first held in short-term memory. Gough and Rumelhart view this information as finely differentiated as series of lines, curves, and angles, while Goodman mentions key letters, e.g., beginning consonants, as at least one type of initially-stored graphic input.

In Rumelhart's model, a feature-extraction process sorts critical from noncritical letter features and deposits critical features into a pattern synthesizer. This pattern synthesizer appears to be a form of short-term memory. For Gough, the parallel process converts letter features directly into letters which are stored in the short-term character register. According to Gough, an abstract phonemic counterpart is then attached to the letter representation and stored on a phonemic tape. The phonemic representations are subsequently acted on by a word-recognition process and stored in what Gough calls primary or working memory—apparently, another form of short-term memory.

Clustering of graphemic information proceeds linearly up to the word level in Gough's model. Goodman, on the other hand, suggests that the original perceptual image of graphemic cues is formed and stored in short-term memory until a search in long-term memory yields related phonological, syntactic, and semantic cues. These latter cues are then transferred to short-term memory for comparison with the perceptual image.

Rumelhart's model draws together many levels of information in a multiple-process, pattern-synthesis stage. Besides critical features of letters previously mentioned, other levels of information include orthographic (letter cluster), lexical (word), syntactic, and semantic cues, all of which are processed to produce "the most probable interpretation" or meaning. Processing at any level of information can effect processing at any other level. Compared to Gough's and Goodman's models, the Rumelhart model is most explicit about disallowing use of extraneous information. For example, in the sentence "Birds of a feather flock together," a reader who had encountered the slogan in his or her previous experience might be able to achieve "the most probable interpretation" by utilizing syntactic and semantic levels of information after relying on lower levels of information for the first three words. Such a leap to higher levels of information would not be possible in Gough's model; and while Goodman would believe such a leap possible, his possible sequences of processes are not so well-defined as in the Rumelhart model.

Word representations from Gough's primary memory are acted on by a sentence-comprehension process that applies syntactic or semantic rules to these words. Newly-understood sentences are then sent to TPWSGWTAU (The Place Where Sentences Go When They Are Understood), or long-term memory—what Gough calls secondary memory. Unlike Rumelhart's view, then, the comprehension process for Gough is isolated and left to the very end of the sequence of reading processes.

Goodman's conversion to meaning, or comprehension, the heart of his model, results in elaborate hypothesis-testing processes. Following comparison of the perceptual image in short-term memory with identified cues from long-term memory, a check for a match between the two is made. If a match is made, then information in some unspecified form is sent to medium-term memory, where another test for appropriateness against previously-developed syntactic and semantic choices occurs. Presumably, these choices are available from prior context. If the choices match, comprehension follows. If choices do not match, other options are available; for example, a reader may regress to the beginning and start the entire sequence again.

Within these models, it is possible to identify common variables that a reader's perception of the learning environment may influence. These include the size of the information units that are processed, the number of units processed per unit of time, and the storage capacity of short-term memory. Simon (1969) examines these variables and defines a "chunk" as a familiar unit of information. He not only shows that from two to seven chunks may be held in short-term memory, but also that it takes five seconds to move a chunk from short-term memory to long-term memory. This time can be lowered, moreover, by increasing the meaningfulness of the chunks. It seems reasonable to hypothesize that a less than optimal environment in the form of inappropriately high levels of competitiveness, friction, and difficulty will cause errors in the input of chunks to short-term memory, reduction in the number of short-term memory and long-term memory items, and slowed processing times. For example, readers' perception of friction in their learning environment may so distract them as to adversely affect their visual-system activities at the very first stage of reading. Because four fixations take just one second, according to Gough, the readers' processing of the lines, curves, and angles may be either incomplete or incorrect. Readers may not garner all the information available for short-term memory, or may not garner sufficient information to fill short-term memory, but incorrectly perceive some of the graphemic input. The fact that readers have not correctly or completely processed the lines, curves, and angles is not as critical in top-down models because, in these models, readers have access to other sources of information, such as world knowledge. Furthermore, the possiblity of regression is not only allowed, but encouraged, in top-down models, thereby providing readers with the opportunity to gather up missed cues.

A reader who is distracted by ineffective class morale in early stages of Gough's version of the reading process will be more hampered in his progress toward comprehension than one who encounters interference in later stages. Because no backtracking occurs in Gough's model, a reader who is closer to the final stage of processing before things go wrong will be more likely to derive some approximation of the author's intended meaning.

In top-down models, as mentioned earlier, even if there are errors in initial processing of chunks of information, they are not as critical as they might be in a bottom-up model for a number of reasons. The regressive and interactive features of Goodman's and Rumelhart's models, respectively, as well as the possibility of relying on more sources of information than initial visual ones alone, should lead to fewer errors based on incorrect or incomplete perception of visual cues. Because top-down models do depend on other sources of information, these other sources are all somewhat susceptible to the influences of learner perceptions of the social environment.

In either Goodman's or Rumelhart's models, comprehension is virtually impossible without input from the reader. For example, hypothesis-testing is a crucial part of Goodman's model that requires readers to make guesses about forthcoming content. Students who are preoccupied by social-psychological problems in their classrooms may be less likely to play the "psycholinguistic guessing game," as Goodman terms this risk-taking application of his model. Likewise, students who are troubled by perceptions of competition, friction, or difficulty in their classroom environments may be unable to synthesize at the most crucial stage of Rumelhart's model.

In conjunction with the above, in these models, readers must bring prior knowledge and expectations of the world around them to their understanding of the printed page. If readers are unduly concerned by psychologically threatening aspects of their environment, they may be less likely to utilize this personal knowledge. Because they may fail to make full use of available "scripts" about things they do know, due to environmental distractions, comprehension may be seriously hampered.

In Goodman's and Rumelhart's models, a reader would, under such conditions, be less vulnerable at any one stage of processing. That is, meaning cues would be accessible from a variety of sources, and would not be blocked by inadequate decoding alone. The time necessary for a given amount of comprehension, however, might be greatly extended by perception of a frustrating environment.

Implications for Research and for the Teaching of Comprehension

The three model-builders selected for comparison implicitly incorporate variables of information-unit size, number processed per unit of time, and short-term memory-storage capacity, although each does so in different ways. Within each model, these variables may be found to be related to the influence of the social-psychological variables in the learning environment perceived by students. As mentioned earlier, student perceptions which index the environment accurately predict gains on learning-outcome measures, for example, reading comprehension scores. This relationship between student perceptions and reading comprehension needs to be more closely explored in terms of actual events in the classroom.

Based upon Talmage and Walberg's (1978) elementary school findings that Competitiveness, Difficulty, and Friction are negatively correlated with gains on a standardized reading test, the following questions may be raised: Where in a reading class situation might students perceive competitiveness, difficulty, and friction, and how might these perceptions be related to reading comprehension as hypothesized by the models?

Some reading situations may prove more conducive to the perceptions of competitiveness and friction than others. For example, with regard to grouping, students involved in cross-class or interclass grouping for reading would find themselves with others of similar reading ability, as more teachers would be available to accommodate the range of reading levels commonly found at any one grade. Such a situation might lend itself to a less competitive atmosphere than the more commonly-found intraclass grouping situation might. Experiments along this line would be interesting to conduct.

Children may also find themselves in more or less competitive environments, depending upon the reading approach used. Children in an individualized reading program, where children work at their own pace on materials specifically tailored for their needs—at least, in the ideal case—may be unaware of other children's reading successes or failures. Those children in a basal reading program, where children are grouped according to their reading ability, may be acutely aware of others' reading successes or failures. Thus, a more likely breeding ground for competition is established.

Other subtle aspects of the classroom environment—what Jackson (1968) terms "the hidden curriculum"—may serve to promote student perceptions of friction. Teachers foster certain ideals by displaying in the classroom only those language-experience stories or book reports that conform to their own expectations, while other work not so favored is returned. A teacher may inadvertently create friction in a classroom by this kind of favoritism. Feelings of friction among peers might also be generated when poor readers are continually corrected in their oral reading by good readers. And, as demonstrated earlier with regard to competition, the reading program itself may engender an environment of friction. For example, reading programs which emphasize only phonics instruction will reward those students who can sound out words. Those students who must rely or who best rely on other strategies when reading, such as making use of context or prior knowledge, will, in such a program, garner less recognition and praise from the teacher. Under these instructional conditions, students who would in another setting normally be called good readers are less favored than students who are comfortable with the phonics strategy. Student perceptions of competitiveness and friction in certain classrooms may, therefore, cause feelings of frustration and impaired reading comprehension.

Reading situations may also foster student perceptions of difficulty. Recent literature provides numerous examples of mismatches which may lead to perceptions of difficulty in the classroom. One such mismatch exists when teachers fail to match instruction to student needs. Carroll (1971) documented the wide variability of reading grade levels from fourth graders to college sophomores and noted that the test performance of the top one percent of the fourth graders was identical to the bottom fifth percentile of the college sophomores. Teachers who do not adapt to such variability will be likely to make unreasonable demands of their low readers, while not demanding enough of their accelerated readers. In this way, reading instruction may be perceived by students either as too difficult or not difficult enough. The statistical evidence gathered thus far suggests that elementary reading classes may be too difficult, on the average, but that high school classes in most subjects are not challenging enough (Haertel, Walberg, & Haertel, 1979, Tables 2 and 3).

In the same vein, research findings on teacher questioning practices are pertinent. Anderson, Spiro, and Montague (1977) favor the use of questioning to extend comprehension; however, Guszak (1968) and Bartolome (1969) both note the prevalence of teachers' questioning at the lower, literal levels. Such practices may stifle the critical thinking abilities of accelerated students.

Goldberg (1973) observed teachers who were stressing comprehension in their reading programs, while students were yet decoding. Teachers were again making inappropriate demands of their students—at least, Gough's view of reading is consistent with this assumption. Guthrie and Tyler's evidence (1978) that indicates that poor readers at all grade levels—but especially at primary grade levels—are hampered by inadequate decoding or word recognition skills also supports Gough here. Insufficient decoding skills are not as great an obstacle to a reader's understanding for the other two model-builders mainly because, for them, routes to meaning are numerous. Poor readers, even in the primary grades, would be expected to show as many comprehension-related as decoding-related deficits.

Not only mismatches of instruction and student needs may occur. Another kind of mismatch leading to perception of difficulty may occur between students' language and experience and textual language and content. Ruddell (1965), among others, showed that comprehension was impeded when syntactic patterns of the text did not match those of children. Unfamiliar language may be perceived to be more confusing and difficult than familiar language patterns. Likewise, reading comprehension has been shown to be enhanced when students possess prior knowledge about the content of the materials read (Gordon, Hansen, & Pearson, 1978). Thus, student perceptions of difficulty caused by mismatching of instruction or materials to readers' needs can easily be linked to reading comprehension.

Other scales not yet studied with regard to the elementary classroom may contribute to perspectives on reading comprehension. For example, student perception of Goal Direction may turn out to correlate positively with reading outcome measures at the elementary as well as the secondary level. A study by the American Institute for Research (Bowers, 1974) found that outstanding reading programs appeared to share many components associated with goal direction:

> ...academic objectives that were clearly stated, that were broken into smaller units, and that gave evidence of careful planning...highly structured teaching directly relevant to the objectives providing both feedback and diagnostic information. (p. 56)

Future research in elementary school reading classes testing the negative influence of Apathy, Friction, Cliqueness, Disorganization, and Favoritism and the positive influence of Democracy, Material Environment, Cohesiveness, and Satisfaction on learning should be interesting and valuable in extending our general knowledge of how social-psychological perceptions influence learning and our specific understanding of the teaching and learning of reading comprehension.

References

ANDERSON, R.C., SPIRO, R.J., & MONTAGUE, W.E. (Eds.). *Schooling and the acquisition of knowledge.* Hillsdale, N.J.: Erlbaum, 1977.

BARTOLOME, P.I. Teachers' objectives and questions in primary reading. *The Reading Teacher*, 1969, *23*, 27-33.

BOWERS, J.E. *Final report identifying, validating, and multimedia packaging of effective reading programs.* Palo Alto, Calif.: American Institute for Research, 1974.

CAMPBELL, W.J. (Ed.). *Scholars in context: The effects of environments on learning.* Sydney: Wiley, 1970.

CARROLL, J.B. Development of native language skills beyond the early years. In C.E. Reed (Ed.), *The learning of language.* New York: Appleton-Century-Crofts, 1971.

GIBSON, E., & LEVIN, H. *The psychology of reading.* Cambridge, Mass.: M.I.T. Press, 1975.

GOLDBERG, M. *The effects of various approaches to beginning reading* (Final Report, Beginning Reading Project). New York: Teachers College, Columbia University, 1973.

GOODMAN, K.S. Behind the eye: What happens in reading. In H. Singer & R.B. Ruddell (Eds.), *Theoretical models and processes of reading.* Newark, Del.: International Reading Association, 1970.

GORDON, C., HANSEN, J., & PEARSON, P.D. *Effect of background knowledge on silent reading comprehension.* Paper presented at the meeting of the American Educational Research Association, Toronto, Ont., April 1978.

GOUGH, P. One second of reading. In J.F. Kavanaugh & I.G. Mattingly (Eds.), *Language by ear and by eye.* Cambridge, Mass.: M.I.T. Press, 1972.

GUSZAK, F. *Questioning strategies of elementary teachers in relation to comprehension.* Paper presented at the meeting of the International Reading Association, Boston, 1968.

GUTHRIE, J.T., & TYLER, S.J. Cognition and instruction of poor readers. *Journal of Reading Behavior*, 1978, *10*(1), 57-78.

HAERETEL, G.D., WALBERG, H.J., & HAERTEL, E.H. *Social-psychological environments and learning: A quantitative synthesis.* Paper presented at the meeting of the American Educational Research Association, San Francisco, April 1979.

INSEL, P.M., & MOOS, R.H. Psychological environments: Expanding the scope of human ecology. *American Psychologist*, 1974, *29*, 179-188.

JACKSON, P.W. *Life in classrooms.* New York: Holt, Rinehart & Winston, 1968.

KAHN, S.B., & WEISS, J. The teaching of affective responses. In R.M.W. Travers (Ed.), *Second handbook of research on teaching.* Chicago: Rand McNally, 1973.

MAJORIBANKS, K. (Ed.). *Environments for learning.* London: National Foundation for Educational Research, 1974.

RANDHAWA, B.S., & FU, L.L.W. Assessment and affect of some classroom environment variables. *Review of Educational Research*, 1973, *43*, 303-321.

RUDDELL, R.B. Effect of the similarity of oral and written patterns of language structure on reading comprehension. *Elementary English*, 1965, *42*, 403-410.

RUMELHART, D. *Toward an interactive model of reading* (Tech. Rep. No. 56). San Diego: University of California, Center for Human Information Processing, July 1975.

SHULMAN, L.S., & TAMIR, P. Research on teaching in the natural sciences. In R.M.W. Travers (Ed.), *Second handbook of research on teaching.* Chicago: Rand McNally, 1973.

SIMON, H.A. *Sciences of the artificial.* Cambridge, Mass.: M.I.T. Press, 1969.

TALMAGE, H., & WALBERG, H. Naturalistic, decision-oriented evaluation of a district reading program. *Journal of Reading Behavior*, 1978, *10*, 185-195.

WALBERG, H.J. Predicting class learning. *American Educational Research Journal*, 1969, *6*, 529-542.

WALBERG, H.J. Educational process evaluation. In M.W. Apple et al. (Eds.), *Educational evaluation: Analysis and responsibility.* Berkeley, Calif.: McCutchan, 1974. (a)

WALBERG, H.J. (Ed.). *Evaluating educational performance; A sourcebook of methods, instruments, and examples.* Berkeley, Calif.: McCutchan, 1974 (b)

WALBERG, H.J. Psychology of learning environments: Behavioral, structural or perceptual? In L.S. Shulman (Ed.), *Review of research in education.* Itasca, Ill.: F.E. Peacock, 1976.

WALBERG, H.J. (Ed.). *Educational environments and effects: Evaluation, research, and policy.* Chicago: National Society for the Study of Education, 1979.

Practices in Education

Instructional Variables in Reading Comprehension

Joseph R. Jenkins
University of Washington

Darlene Pany
Arizona State University

Our aim in this paper is to summarize the scientific knowledge base which underlies reading comprehension instruction in the middle elementary years. Of particular interest are specific instructional practices rather than comprehensive reading curricula as represented by basal reading programs. Little can be said about either the relative or absolute effects of various reading curricula since virtually no research exists on the subject (Jenkins & Pany, in press). That there is so little information about the effects of reading curricula is remarkable given the overwhelming reliance on these programs by schools. Even if data were available on the effectiveness of various reading curricula it would be difficult to determine which aspects of the programs were functional and which were frivolous, since the programs are comprehensive and each contains a broad collection of instructional practices. Relatively more research has been conducted on a host of more circumscribed instructional variables.

The decision to focus on studies conducted with children in the middle elementary grades was based on several considerations. Below grade three, reading instruction typically emphasizes word decoding rather than comprehension, a practice that is not without its critics (Smith, 1973). Beyond grade eight, classroom instruction becomes increasingly content oriented, with less emphasis given to reading process. It is in the middle elementary grades that schools explicitly admit to teaching reading comprehension (Mason & Boggs, Note 1). By confining our review to studies conducted with children as opposed to those with more mature readers, we do not mean to imply that studies with the latter group are without relevance, only that we are primarily interested in instructional factors which affect the *development* of the ability to comprehend written discourse.

For the purpose of this chapter, we liberally define reading comprehension and reading comprehension instruction. We accept as evidence of comprehension such varied performances as answering passage-dependent questions, retelling facts or ideas stated in a reading selection, completing cloze tests, and orally by reading passages for miscue analysis. Likewise, we take a broad view of reading comprehension instruction to include deliberate attempts by a researcher to modify children's comprehension or memory for prose. In our sense of the term instructional practices for reading comprehension are those factors *external* to the text which can be manipulated to influence comprehension and memory for it. Thus, such variables as advance organizers, pictures, or paragraph heads can be thought of as instructional variables. Teaching children skills related to rapid decoding, vocabulary, listening, cloze performance, and imagery also qualify for inclusion since instruction focuses on factors external to the text. Not all variables that influence comprehension qualify as instructional variables, however. For example, story plot (Thorndyke, 1977), text organization (Meyer, 1975), and syntactic structures (Chomsky, 1972) are factors which influence the comprehensibility of prose, but they are not instructional variables in the sense in which we use this term. Passage characteristics such as those cited above definitely can affect a reader's acquisition of an author's message and can legitimately be considered instructional variables with respect to this intended message. However, since they are characteristics of a particular passage they cannot be manipulated without changing the passage itself. We do not consider the modification of a passage to make it more comprehensible to be an instance of *teaching* reading comprehension.

With reading comprehension and its instruction defined in this way, there is considerable diversity in the studies which qualify as instructional research on reading comprehension. The diversity is a serious obstacle for anyone attempting to organize and comment on the present status of knowledge in the teaching of reading comprehension. A good theory of reading would certainly help to identify promising instructional interventions and to aid in the understanding and classification of variables affecting comprehension, but such a theory has yet to emerge. Without a theory and in search of a framework in which to organize a wide array of instructional variables, we noted that we could identify lines of research which appeared to focus on different aspects of reading comprehension. The variables chosen for study, to some extent, reflect what an investigator believed to be the source of reading comprehension failure. We were able to identify three general sources of comprehension failure which could serve as a basis for grouping instructional research on this problem.

Causes of Reading Comprehension Failure

One line of instructional research emanates from a "top down" view of the reading process. This view highlights the individual's use of existing knowledge structures to interpret and organize prose. Individuals might fail to comprehend because they lack the appropriate abstract knowledge structures or schemata needed for making sense out of text (Anderson, 1977; Spiro, 1978). In some instances the individual may even possess appropriate schemata but for some reason the text does not activate these structures. In either case the individual is unable to make the necessary cognitive contributions required for producing the "click" of comprehension (Brown, 1968). Classic illustrations of these two situations have been provided by Bransford and Johnson (1972). From the standpoint of instruction, a variable such as an advance organizer reflects a top-down orientation. Organizers are hypothesized to enhance comprehension through their influence on cognitive structure.

In contrast to this top-down orientation, other analyses of comprehension emphasize "bottom-up" or text-driven processing (Spiro, 1979). Concern is directed more toward the linguistic demands of the text than toward cognitive structure variables and background knowledge of the person reading the text. According to this "bottom-up" analysis, comprehension problems are likely to result from passage-related variables such as complex or unfamiliar vocabulary, sentence syntax, and text organization or from passages that require special thinking skills (for example, syllogistic reasoning). Accordingly, comprehension instruction attempts to provide the individual with the essential skills needed to simplify the linguistic input of text. Examples of comprehension instruction that emanate from a bottom-up orientation include teaching children word meanings, to decode rapidly, to interpret complex sentences, and to apply reasoning strategies.

Whereas these two views of reading comprehension emphasize schema (top-down) and text-driven (bottom-up) processing, the third analysis focuses on attentional factors and levels of processing (Craik & Lockhart, 1972). Of interest are comprehension problems that result from inattentive, superficial reading or from misdirected attention (e.g. concern with correct word calling rather than meaning getting). Here, the reader is assumed to possess adequate background knowledge along with satisfactory linguistic and reasoning skill, but may not have learned to maintain focus on and self-monitor comprehension, or to employ systematic memory strategies. Instructional research examines variables that curriculum writers may employ to maintain reader attention (e.g. adjunct questions), that readers may use to monitor their own memory or comprehension processes (e.g. imaging), and that teachers may use to shape attentional focus (e.g. corrective feedback that emphasizes meaning vs. sound symbol relations).

We have categorized instructional research on reading comprehension into these three classifications: background knowledge, or schemata, linguistic/reasoning skills, and attentional/processing factors. Admittedly, not all the instructional practices we examined fit neatly into these classifications and one can take exception to our classification of particular variables. For that matter, we take exception to our own classification at times, if that provides any comfort. Nevertheless, these categories should provide a framework for grouping some very different instructional variables, and should help prompt speculations as to what a researcher believes is the source of failure to comprehend.

Instructional Variables Addressing Schema-Related Problems

Students may fail to comprehend a passage because they lack the relevant background knowledge or schemata needed for constructing meaning from the textual recipe. Reading instruction that focuses on improving language abilities or on increasing semantic processing will not resolve the comprehension problem since the problem results from a reader's inability to make the needed cognitive contributions to a text rather than from language or attentional inadequacies.

Does background knowledge really affect comprehension in normal reading situations? It is noteworthy that the most convincing demonstrations of the importance of background knowledge have occurred with contrived, and often ambiguous, reading passages (Bransford & Johnson, 1972). However, a recent study by Pearson, Hansen, and Gordon (1979) suggests that differences in background knowledge may indeed account for a significant portion of the variance in comprehension performances in normal reading situations, in this case with passages taken from a second grade reader. Pearson et al. selected children who were comparable in intelligence and in "reading comprehension ability," but who varied in background knowledge for a particular topic. These students' comprehension of a passage on that topic appeared to be a function of their prior knowledge of the topic. Notably the effects of background knowledge were more pronounced on "scriptually implicit" questions (those that require an integration of textual and background data) than for "textually explicit" questions (those that require responses directly from the text). Pearson et al. interpret their findings as supportive of the notion that comprehension involves integrating new information into pre-established schemata and that such integration is seriously impaired for students with poorly developed schemata. Their conclusion would be strengthened considerably if the experimenters had employed a "test only" control. It could be argued that comprehension differences between high and low schema groups would have existed even if students had not read the experimental passage, but had merely been tested.

Several studies that attempted to facilitate reading comprehension have employed variables intended to establish, modify, or activate background knowledge or schemata. We include research on advance organizers, pictures, and certain textual adjuncts such as titles and paragraph headings. We interpret advance organizers and pictures primarily as attempts to provide the reader with background knowledge or relevant schemata. Titles and paragraph heads (and sometimes advance organizers and pictures) may be thought of as mechanisms for helping the reader identify an already available schematic framework with which to integrate textual information. Research on these variables is examined in this section.

Advance Organizers

Ausubel (1963) has formulated one instructional procedure for enhancing reading comprehension that gives special consideration to the learner's background knowledge. He conceptualized the advance organizer as a "... bridge... between what the learner already knows and what he needs to know in order to learn new subject matter effectively" (Ausubel, 1978, p. 253). An advance organizer consists of introductory material which is related to already existing ideas in the learner's head and which is written at a higher level of abstraction, generality, and inclusiveness than the learning passage itself.

While numerous researchers have investigated the effectiveness of advance organizers, their findings have been largely inconsistent. More often than not, however, no advantage was found for groups who received an advance organizer (Barnes & Clawson, 1975). Several explanations could account for the apparent failure of advance organizer interventions. It is possible that organizers did not assist readers to relate new information to existing knowledge. This could occur if the organizer failed to tie together the two sets of information, or if the reader did not possess the presumed prior knowledge to which the organizer was addressed. Another possibility is that readers in these experiments already possessed readily available schemata for the learning passages, and for them the advance organizer treatment was superfluous. Thus, the performance of an organizer group would not exceed that of a control group. Whatever the explanation, advance organizers, at least in their present state of definition, do not represent a high success intervention for enhancing comprehension.

Titles and Paragraph Heads

Under certain conditions, for example, when asked to read highly ambiguous passages, mature readers' comprehension can be considerably influenced if they are given thematic story titles (Bransford & McCarrell, 1975). Presumably the title offers a relevant framework with which to interpret or disambiguate the passages.

Doctorow, Wittrock, and Marks (1978) extended the use of titles to unambiguous prose passages. Sixth grade students read passages that were accompanied by headings, which were high-frequency synonyms for words central to each paragraph's topic. These students answered 43 percent more comprehension questions correctly than did students who read the passages without the aid of the headings. In interpreting their findings, Doctorow et al. propose that paragraph headings function as retrieval cues which assist the reader in locating relevant memories, or background knowledge. Presumably, the retrieval of these memories allows the reader to construct meaningful elaborations for the textual information. In turn, these elaborations enhance comprehension and memory for the text.

These findings of Doctorow et al. are inconsistent with results of several other investigations. In other studies of children (Landry, 1967; Snavely, 1962), adolescents (Cole, 1977), and adults (Christensen & Stordahl, 1955), no advantage was found for passages with paragraph headings. It is difficult to reconcile these conflicting results. Fortunately, the best designed experiment of the group is the one which reported comprehension facilitation from the paragraph headings. It is also the only study which carefully defined the procedures for creating the thematic paragraph headings. The sizable effects reported by Doctorow et al. warrant further research in this area.

Pictures

Pictures enhance comprehension in both reading and listening tasks (Levin & Lesgold, in press; Schallert, in press). Despite some evidence to the contrary (Samuels, 1970), the data showing facilitation by pictures are remarkably consistent. In reviewing the literature on picture use, Schallert concludes that pictures will be helpful:

> ...when they illustrate information central to the text, when they represent new content which is important to the overall message, and when they depict structural relationships mentioned in the text. In addition, pictures seem to have a specific effect which is localized mainly to illustrate information and which amounts to more than a second rehearsal of the text [in press].

Moreover, the effects of pictures are reasonably robust, with facilitation observed across a wide age range, large social class differences, varying intelligence levels, varying lengths of passages, verbatim and paraphrase test items, and immediate and delayed tests (Levin & Lesgold, in press).

Do pictures enhance comprehension because they alter schema? We feel it is only partially correct to classify pictures as a background knowledge variable which affects comprehension. In some circumstances, pictures definitely provide a framework for interpreting textual information. Bransford and Johnson's (1972) Modern Day Romeo experiment, in which the text is incomprehensible without an accompanying picture, represents the most compelling demonstration of a picture functioning as a schematic, contextual variable. In other circumstances, pictures may enhance comprehension because they provide different and/or better information than text. Kolers (1973) and Schallert (in press) both hypothesize that compared to words, illustrations can better depict spatial and structural information. Finally, in those circumstances where pictures have enhanced comprehension despite the fact that the information to be derived from the picture was explicity stated in the text, the pictures may have modified the reader's attention and semantic processing of the text. For example, pictures may induce the reader to actively elaborate the text with mental imagery.

Regardless of the reason for the facilitative effects of pictures, they appear to represent an instructional variable that a curriculum developer or teacher could use to promote reading comprehension. We would argue, however, that pictures have rather circumscribed effects. Like advance organizers, titles, and subheadings, they are an instructional variable appropriate to situations where the student is reading for new information as opposed to learning to read. There is no evidence that teachers can use pictures to make their students better comprehenders in any general sense. Providing children with texts plus pictures is not likely to affect their subsequent success in comprehending other text without pictures. On the other hand, if background knowledge is as important to reading comprehension as it appears to be (Anderson, 1977), then variables such as pictures and paragraph headings will impact, at least indirectly, on reading comprehension ability because the reader will have acquired broader and deeper background knowledge, having read materials which were accompanied by these aids.

Most background knowledge variables, such as organizers, titles, and pictures, do not represent skills which would be taught. Rather, they are adjuncts to specific passages that are supplied by authors and editors to help readers integrate the passage information with their existing knowledge. A different kind of background knowledge variable is represented by rules or generalizations which affect the reader's ability to interpret information from a variety of passages. A study by Clements, Stevens, Kameenui, and Carnine (Note 2) illustrates this kind of prior knowledge variable. In their study, they taught two groups of students to identify and interpret motives of story characters. Stories presented either an apparent motivation of a character or both an apparent and true character motivation which had to be inferred from explicit information given in the stories. In a Rule and Questioning treatment, students were told: "There may be more than one reason why someone does or says something." Students then orally read a story, heard the rule again, and proceeded to answer a series of questions designed to lead them to infer the true motive of the character in the story. In a Corrective Feedback treatment, students read a story and answered a question about a character's true motive. If students' answers were incorrect or partially correct, the experimenter modeled the correct answer, but did not tell the rationale or derivation of the answer. Subsequently students were given novel passages and requested to identify character motives. The two training groups' performance was equivalent, and superior to that of an untrained group. Apparently students were able to induce the generalization about real and apparent motives from mere corrections and did not require the explicit statement of a rule, nor the explanation of how the rule applied to specific stories. This study suggests that certain general rules can be taught which help students interpret different aspects of text. These rules could be considered a type of high level advance organizer or schemata which are applicable to many different texts.

Instructional Variables Addressing Linguistic/Reasoning Problems

While the reader's prior knowledge certainly plays an important role in what will be learned from text, it is not the only variable. It stands to reason that aspects of the text itself, the reader's ability to deal with various linguistic structures, and the ability to reason about and interpret different kinds of messages all can affect the degree to which a text will be understood and remembered. In this section we will review attempts to provide the reader with skills needed to simplify difficult aspects of text. Included are rapid decoding, vocabulary, cloze, organizational strategies, specific subskills instruction, and auding-reading.

Rapid Decoding

Most reading authorities agree that some level of decoding proficiency is necessary for adequate reading comprehension. In his text on the teaching of reading, Harris (1970) states, "Some very slow readers do poorly in comprehension because their many repetitions and hesitations break up the continuity of thought" (p. 447). Consistent with the notion that decoding speed is a factor which influences reading comprehension, Perfetti and Hogaboam (1975) found that good and poor comprehenders differ in the speed with which they decode single words. Based on these data, Perfetti (1977) proposed a shared capacity or "bottleneck" hypothesis to account for the relationship between decoding speed and comprehension. The basic notion is that individuals possess limited processing space, and that decoding and comprehension are separate but interrelated tasks that both require that space. The more processing space consumed by decoding, the less processing space available for comprehension; thus, inefficient decoding can detract from comprehension.

LaBerge and Samuels (1974) also argue for the importance of decoding speed or automaticity. In their model, the development of reading skills is marked by an increase in automatic processing of print. Automatic processing which results from advanced familiarity with letters and words is characterized by rapid and accurate response to print, such that the reader does not expend attention on the task of decoding. A proficient reader automatically processes words and their meanings, thus releasing attention for comprehending.

The data base for the presumed influence of decoding on comprehension is essentially correlational and, as Perfetti (1977) has indicated, the basis for asserting the causal relationship between decoding rates and comprehension is still conceptual, not empirical. That rapid automatic decoding should facilitate comprehension has enormous face validity and this notion gains support from data showing that good comprehenders tend also to be rapid decoders. But, it does not necessarily follow that helping children to become rapid decoders will help their comprehension. Indeed, little prior research has been conducted with poor readers on the effects of training in rapid decoding. Two studies (Dahl, 1979; Samuels, Dahl, & Archwamety, 1974) attempted to examine this issue experimentally. Their results indicated that groups of students who had received speeded isolated word training performed no better on comprehension tests than did untrained students. However, in neither of these studies did word drill produce effects on speed of word recognition; thus, failure to observe transfer effects to comprehension should have come as no surprise.

A stronger test of the rapid decoding hypothesis was made in two experiments by Fleisher, Jenkins, and Pany (1979). Fourth and fifth grade students were classified as good and poor readers according to their reading comprehension test scores. One-half of the poor readers received rapid decoding training on all words from a test passage. Word drill was continued until these poor readers were able to read all words that would appear in the test passage at a rate equivalent to that achieved by good readers. Next, test passages were presented to good readers, to poor readers who had received rapid decoding training, and to poor readers who had received no training. After reading the passages the students were given several comprehension measures including questions, story retell, and cloze. Although there were minor procedural differences in the two experiments, in both cases the results indicated that poor readers who had received rapid decoding training performed no better than their nontrained poor reader counterparts. Both groups of poor readers performed significantly worse than the good reader group on the comprehension measures, even though the trained poor readers read the passages more fluently than their untrained counterparts. However, despite the training in rapid decoding, poor readers still read significantly less fluently in context than the good reader control group. These results suggest that single word training does not automatically transfer to comprehension, and does not produce high levels of fluency in context.

Dahl (1979) reported more optimistic results in an extremely well designed study testing the rapid decoding hypothesis. Her study deserves a high mark on the dimension of ecological validity in that it was conducted in the context of a fourth grade classroom on a daily basis throughout an entire school year. To develop fluency, experimental students practiced reading 100-word passages. They reread each passage until they could complete it in one minute. Whenever a student achieved this goal on one practice passage, that passage was replaced by another. At the end of the year, these children were compared with control children who had engaged in "regular" reading instruction for an equivalent length of time. Dahl's results indicated that children who were given repeated reading practice (to develop automaticity) exhibited superior performance on cloze measures, on errors during oral reading, on reading rate, and on isolated word recognition. Also, it appeared that these children achieved significantly higher scores on the Gates-MacGinitie reading test (another comprehension measure), although it is impossible to be certain of this, since Dahl's report did not include statistical details.

Dahl's findings are clearly at variance with those of Fleisher et al. There are several possible explanations for the discrepancies. Whereas the Fleisher et al. training procedure emphasized single-word decoding, Dahl's repeated reading procedure provided extensive practice on speeded reading in context. This difference along with the duration of Dahl's treatment may have enabled her students to improve not only their fluency but also their organizational skills, for example, they could better segment complex sentences into meaningful units. Moreover, it is possible that students need to attain a level of fluency and maintain it for some period before they can take advantage of their newly gained processing resources and apply them to comprehension work. By themselves, Dahl's findings provide sufficient grounds for remaining attentive to decoding speed and accuracy when considering instructional interventions to improve reading comprehension.

Vocabulary

The importance of word meaning knowledge to reading comprehension would seem to be self-evident. According to Spache (1966), "Understanding the vocabulary is second only to the factor of reasoning in the process of comprehension, and some writers would say it is even more important than reasoning.... It is sufficient to say that comprehension is significantly promoted by attention to vocabulary growth" (p.78). While the research on vocabulary is enormous (Dale, Razik, & Petty, 1973), it is largely descriptive. Unfortunately, there are relatively few studies which directly document the effects of vocabulary instruction on reading comprehension.

In studies that have examined the effects of vocabulary instruction on general measures of reading comprehension (i.e., standardized tests), the findings have not been encouraging (e.g., Currie, 1963; Hafner, 1965; Lieberman, 1967; Otterman, 1955). Only Currie reported a significant comprehension effect. Students who completed exercises in classifying words according to similarities of structure and meaning had higher comprehension scores than did control students. Teaching students vocabulary through context aids (Hafner, 1965), word-roots and prefix meanings (Otterman, 1955), or concept development (Lieberman, 1967) appeared not to affect reading comprehension as measured by standardized tests.

When researchers have studied the effects of vocabulary training that focused on specific words and specific passages, the results have been equally disappointing. In a series of experiments, Pany and Jenkins (1978), Jenkins, Pany, and Schreck (1978), and Pany (1978) evaluated the effects of several vocabulary instruction procedures on a variety of measures. The instructional procedures included drilling on synonyms, telling word meanings in the context of oral reading, relating words to common experiences, and providing practice in applying word meanings. Based on a number of vocabulary measures, synonym drill was consistently the most effective instructional procedure. The differential effects of the various instructional procedures on vocabulary acquisition were generally mirrored on measures of *sentence* comprehension (Jenkins et al., 1978; Pany, 1978). In contrast, vocabulary training produced no discernable effects on the comprehension of *passages* containing the words, which had been taught whether comprehension was measured by questions, cloze, or retell.

These results were partially supported in research conducted by Kameenui and Carnine (Note 3) who compared reading comprehension effects from two vocabulary training procedures. For one group, training consisted of practice on synonym meanings and word application. A second group received this training plus integration instruction in which the student was stopped during oral reading and asked to generate meaning for those sentences which contained a newly learned vocabulary word. These training groups then read a test passage which contained the new vocabulary. Their comprehension was compared with a control group which read the same test passage. On literal comprehension questions there were no differences among groups, but on inferential questions the control group which had not received vocabulary training performed significantly worse than the training groups, whose performance was equivalent.

In summary, while a variety of methods are capable of teaching vocabulary and of affecting comprehension at the sentence level it is not so easy to affect overall reading comprehension. The effects of vocabulary knowledge on reading comprehension seem to be far more subtle than many reading educators had imagined.

Cloze

The cloze task has become a common measure of reading comprehension. In this task students are asked to supply words that have been deleted from a passage. The number of correctly supplied words serves as a reading comprehension score. Cloze presumably taps a student's ability to use available linguistic information to construct meaning for a passage. The rationale for using cloze as an instructional activity is that students become sensitive to the semantic and syntactic cues in text and learn to "test" the sensibility of the text produced from their word substitutions.

What are the effects of providing students with practice on cloze tasks? In the studies that have examined this procedure, few have noted improvements in reading comprehension. Several of these studies (Blumenfield & Miller, 1966; Guice, 1969; Heitzman & Bloomer, 1967) used older subjects and thus fall outside the scope of this report. However, two studies which used sixth and seventh grade youngsters (Louthan, 1965; Schneyer, 1965) also report nonsignificant comprehension differences between cloze-practice and control groups. Neither experiment used a cloze comprehension measure. In contrast to these findings, Kennedy and Weener (1973) reported enhanced comprehension after practice with cloze. They gave third grade remedial students cloze exercises which were presented either through a reading or listening mode. Compared to a non-instructed group, the cloze-practice group scored significantly higher on the Durrell Listening Test and on reading and listening cloze measures. On the Durrell Reading Test, the cloze-trained group exceeded the performance of all other groups. What is remarkable about Kennedy and Weener's study is that any effects were found, given the limited amount of training. Cloze practice consisted of only 100 items and no more than 1 hour and 40 minutes of instruction.

Samuels et al. (1974) developed an instructional procedure which is distinct from other cloze research. They analyzed the skills involved in completing cloze exercises into seven components, including training in the use of auditory and visual context to predict words which could logically follow in a sentence. Samuels et al. termed this procedure "hypothesis/test training." Groups trained in hypothesis/test consistently outperformed control groups on cloze measures of comprehension (Dahl, 1979; Samuels et al., 197 Sindelar, Note 4). On standardized reading comprehension measures, Dahl reported differences which favored the hypothesis/test group while Sindelar found no significant differences. Samuels et al. and Dahl attributed their findings of improved comprehension to the children's increased automaticity. They believe that the children receiving hypothesis/test training became adept at word recognition using partial rather than whole word cues.

The results of studies employing cloze as an instructional technique, while far from consistent, suggest that certain uses of cloze may benefit reading comprehension. The hypothesis/test procedure or some variation thereof, which involves training on the subskills which presumably contribute to cloze proficiency, may be the key to effective cloze instruction.

Organizational Strategies During Reading

Research on readability indicates that a number of text characteristics influence comprehension (Bormuth, 1966). Several studies have evaluated the effects of syntactical and semantic factors on children's understanding of and memory for text (Bormuth, Manning, Carr, & Pearson, 1970; Lesgold, 1974; Pearson 1974-1975). However, research on *training* children to understand syntactical structures is sparse.

In one study, Weaver (1979) attempted to influence reading comprehension by teaching children to encode text in meaningful chunks. She trained average and above average third grade readers to unscramble sentence anagrams as a means of teaching children to chunk meaningful units within sentences. Her results were impressive; after an average of 3½ hours of practice on sentence anagrams, experimental subjects significantly outperformed untrained controls on speed and accuracy of unscrambling anagrams, on cloze paragraph performance, and on a prompted verbatim sentence recall test. However, differences were not found on the only standardized reading comprehension test used (MAT Reading Subtest) and on a timed test involving recognition of meaningful sentences.

Another training activity which is thought to facilitate "chunking" of information is practice in sentence combining. In general, sentence-combining exercises are designed to increase students' awareness of larger linguistic units and thus increase language fluency. In two experiments (Fisher, 1973; Hughes, 1975), students were given sets of sentences to combine in writing (e.g., "My brother tore up my book" and "My brother is little" results in "My little brother tore up my book"). After five hours of instruction, fifth, seventh, and ninth grade students outscored controls on the Stanford Paragraph Meaning subtest and a cloze test (Fisher, 1973). In contrast, Hughes (1975) provided sentence combining training for low, middle, and high ability seventh graders. After a minimum of 26 hours of practice, there were no significant differences between experimental and control groups on a cloze test or the Gates-MacGinitie subtests for Speed and Accuracy; but on the Miscue Inventory (Goodman & Burke, 1972), experimental groups outperformed controls on comprehension and grammatical strength. Hughes observed that the greatest gains in reading comprehension from sentence combining occurred in the lower and middle ability groups.

Research on teaching organizational strategies, although limited, holds promise for improving reading comprehension. Activities which enhance "chunking" of information appear to affect students' reading behavior.

Specific Subskill Instruction

While reading researchers debate the subskill vs. holistic nature of the reading process, daily reading instruction in schools is typically approached from the subskill perspective (Jenkins & Pany, in press). Commercial reading materials generally offer activities designed to teach such comprehension skills as "finding the main idea," "identifying the sequence," and "drawing inferences." "Instruction" is essentially a matter of discovery learning through workbook exercises and teacher questioning, rather than through direct rule learning, rule application, or task analytic procedures. Several studies have addressed the teaching of specific comprehension subskills, and these have been based on a task analytic direct instruction approach. For example, Carnine and his associates have used these procedures to teach a number of subskills ranging from simple to complex. These include answering sequence and simple fact questions (Carnine, Prill, & Armstrong, Note 5), using context to derive the meaning of unknown words (Coyle, Kameenui, & Carnine, Note 6), comprehension of reversible passive voice and clause constructions (Kameenui, Carnine, & Maggs, Note 7) and selecting information needed to form conclusions (Woolfson, Kameenui, & Carnine, Note 8).

Another comprehension subskill is "critical reading." Robinson (1964, p. 3) offers a general definition of critical reading as "judgement of the veracity, validity, or worth of what is read, based on sound criteria, or standards developed through previous experience." Wolf, King, and Huck (1968) report scores on the Ohio State University Critical Reading Test which favored groups who for one year received lessons involving reading, discussing, and evaluating printed materials and completing logic worksheets. Nardelli (1957) reported that students who received instruction made significantly greater gains on recognition of propaganda devices, but not on interpreting authors' suggestions or characters' feelings. Students who used workbooks and audio tapes to learn logic rules (Lowerre & Scandura, 1973-74) exhibited significant growth on a test which measured ability to recognize valid inferences, to detect statements incompatible with the premises, and to detect conclusions that were not logically permitted by the premises. In three experiments designed to teach children to contend with determinate and indeterminate syllogisms in text, Katzenmeyer and Van Blaricom (1976) report "... that although the results were statistically significant [on one of three experiments], there were no major improvements in scores due to instruction [on syllogisms]" (p. 5).

In summary, it appears that students can acquire specific comprehension subskills if they receive carefully designed instruction. A question that is yet unanswered is whether mastery of specific subskills affects overall comprehension. None of the studies reported in this section offer evidence that children who receive instruction in a particular skill area demonstrate improved comprehension on noncontrived texts. Do students who learn to comprehend passive voice constructions in specific exercises apply that skill in more natural context? Do students who receive specific subskill training recognize propaganda or passive voice constructions or problem statements when they encounter them in other contexts? A serious void in the comprehension subskill research to date is the failure of researchers to assess transfer effects to other reading comprehension skills and to more general measures of comprehension. While it appears that specific comprehension skills *can* be taught, how this accomplishment relates to more general reading comprehension ability has yet to be established.

Auding-Reading Relations

One basis for designing reading comprehension instructional strategies is provided by the developmental model of auding and reading described in the work of Fries (1963) and Sticht and his colleagues (Sticht, Beck, Hauke, Kleiman, & James, 1974). According to this comprehension model, the same language competencies and cognitive content which permit one to understand spoken messages (auding) are also responsible for the understanding of written messages (reading). Thus once a person can decode print with sufficient automaticity, his/her comprehension of text will be equal to that which would be achieved if the message were processed auditorally.

In one sense, research on the auding-reading relationship falls outside the organizational framework of this chapter. In theory, all instructional variables could be examined in relation to the aud-read model. However, since auding-reading research has largely examined variables which allow the reader to simplify the linguistic input of passage, we are including it in this section.

Sticht et al. (1974) offer four lines of evidence which support predictions from their auding-reading model. First, the ability of children to comprehend spoken messages should exceed their ability to comprehend written messages during the early school years, presumably because beginning readers lack competency in decoding printed language. As decoding skills develop over time, auding and reading abilities will become equal. In general, the existing evidence related to this prediction confirms that auding is initially superior to reading and that the two processes become equally effective sometime around seventh or eighth grade. Second, auding ability will predict reading comprehension, once decoding skills develop. Data relating to this prediction shows that the correlations between these abilities increase from first to fourth grades and remain stable (around +.60) thereafter. Moreover, upper elementary and junior high aged students who are poor readers also perform badly on listening tasks, suggesting a general comprehension deficit rather than one specific to reading (Smiley, Oakley, Worthen, Campione, & Brown 1977; Becker & Lilly, Note 9). Third, if similar language and cognitive competencies are responsible for both auding and reading, then similar rates of information input, whether by printed or spoken language, should yield similar levels of comprehension. Evidence offered in support of this prediction reveals that the maximal rates for silent reading with accurate retention are similar to rates of both auding and speaking (250-300 wds./min.). Fourth, and most crucial for instruction in comprehension, the model predicts that training which improves auding ability will be reflected in improved reading ability at least for those individuals who can decode print. Of twelve studies reviewed by Sticht et al. (1974) in which specific auding abilities were improved through training, ten reported improved reading ability which paralleled the improvement in auding ability. Those studies which reported an auding/reading transfer tended to include such content as "training in listening in order to recall events, ideas, or details . . .training in vocabulary . . .training in listening to predict outcomes or to draw conclusions or inferences . . .[and] training in listening to follow directions" (Sticht et al., p. 88).

Another study bearing on auding/reading transfer was reported by Kennedy and Weener (1973) and previously reviewed in our section on "Cloze." Compared to a noninstructed control group, children who underwent cloze auding training scored significantly higher on the Durrell Listening Comprehension test, on a cloze listening measure, and on a cloze reading measure.

Implications of the developmental model of auding and reading for reading comprehension instruction are noteworthy. Once decoding is mastered, improvements in reading comprehension should be a function of improvements in the reader's language and cognitive competencies. Efforts to improve language and cognitive competencies can occur in the context of reading, but are not necessarily limited to reading activities. It would seem that instructional activities which expand on the child's world knowledge and on his ability to extract meaning from various syntactical forms of discourse would produce reading comprehension improvements. The aud-read model implies, further, that specific language competencies of children should be assessed, so that instruction could be geared to appropriate targets. If the aud-read model is valid, two observations are germane. Some "reading comprehension" problems should conveniently disappear as soon as children achieve proficiency in decoding, at least in children who have an adequate knowledge and language base. Thus, the necessity of producing proficient decoders is obvious. Second, although instruction designed to improve cognition and language abilities may begin early, it should not be expected to have maximum impact on reading comprehension until decoding skills are reasonably proficient.

Instructional Variables Addressing Attention Related Problems

Another source of comprehension failure involves the attention a reader gives to the task. We would expect comprehension to suffer if students processed text superficially, attended to unimportant aspects of the task for example, pronunciation, or failed to self-monitor their understanding and memory for what is read. Instructional variables which seem to address these attention related problems include questions, purpose setting, incentives, and reading strategies.

Questions and Purpose Setting

Establishing a purpose for reading and asking children questions about what they read are two frequently recommended instructional procedures (Weintraub, 1969). In a recent investigation of classroom reading instruction, Durkin (1978) found that teachers devoted the largest time allotment (18 percent) of the reading period to comprehension assessment, that is asking questions about what was read. To our knowledge, there have been no systematic investigations of the effects of teacher questioning on the development of children's comprehension. In contrast, there is an enormous literature on the effects of adjunct questions in text and there is some research on purpose setting. While most of this research has focused on mature readers, e.g., college students, several recent studies have used younger students.

Providing children with purpose statements such as "read for implied meanings" or "read to find out how . . ." before they read a passage does not appear to facilitate comprehension (Ballard, 1965; Pettit, 1971; Snavely, 1962). Similarly, placing questions before a reading passage does not consistently increase comprehension relative to no-question conditions (Fincke, 1968; Landry, 1967). In contrast, interspersed questions, either before or after pertinent passage segments, have sometimes been shown to aid children's subsequent test performance. These effects, however, are by no means consistent across studies. For example, Daugherty (1971) found that interspersed post-questions did not facilitate comprehension, while Swenson and Kulhavy (1974) and Yost, Avila, and Vexler (1977) found that interspersed post-questions strongly facilitated comprehension. Swenson and Kulhavy also reported significantly higher performance on relevant items (posttest information that was questioned during reading) vs. incidental items (posttest information not questioned during reading). Richmond (1976) reports results that fall somewhere in the middle of these conflicting findings. His study examined the generality of question effects across reading selections. Three different passages were examined under two interspersed-question conditions (before and after) and a no-question control. In contrast to Swenson and Kulhavy and Yost et al., Richmond found no overall effects for the question conditions nor for a relevant vs. incidental item comparison, although the statistical analysis was quite powerful (N 600) and questions were carefully constructed. However, Richmond reported significant interactions among passages, questions, and the relevance dimensions. These data illustrate the highly interactive nature of the question and passage variables which might somewhat explain the discrepancies in results obtained by various investigators.

In general, the results of the question/purpose setting studies do not yield clear signals for practitioners. However, since interspersed questions sometimes elicit more attention from the reader, teachers may wish to use them in reading assignments. But even when interpersed questions raise test performance, do they improve children's ability to comprehend? This is the same issue that was raised in the discussion of advance organizers, pictures, and titles. Again, our speculation will have to suffice. On those occasions when questions do affect comprehension and memory for text, they should increase children's knowledge of the world. If world knowledge is important for reading comprehension, then questions should improve, albeit *indirectly*, children's ability to comprehend subsequent reading selections, at least if these selections are related to that niche of world knowledge previously taught.

Whether a history of reading with adjunct questions *directly* affects children's ability to comprehend other selections that do not have question adjuncts is altogether a different question. Perhaps such questions have only passage-specific effects in that they only alter children's reading strategies temporarily, e.g., by increasing semantic processing. Or perhaps, after children have had extensive experience in reading with adjunct questions, their general reading strategies become permanently altered so that they read any text (with or without adjunct questions) with more intense semantic processing.

Incentives

The problem of poor comprehension is nearly always attributed to skill or knowledge deficits. Consequently, problem solutions usually involve teaching low performing students some skill or strategy or providing them with relevant background knowledge. For some children, however, poor comprehension may not be the result of poor instruction but rather lack of motivation. Several studies conducted in school settings with normal reading materials have investigated the effects of contingency management procedures.

Lahey, McNees, and Brown (1973) and Jenkins, Barksdale, and Clinton (1978) report significant improvements in the number of correctly answered comprehension questions when such activity was reinforced by social praise and money. Lovitt and Hansen (1976) and Hansen and Lovitt (1976) have also modified reading comprehension with contingency management procedures. In one study, permission to skip stories in the reading book was contingent upon a prescribed level of comprehension performance. If the children performed beneath the prescribed level, they were required to reread the passages and correct their answers. When these contingencies were in effect, reading comprehension improved.

This set of studies suggests that comprehension performance can be altered through contingency management. Like adjunct questions, incentives have uncertain direct and indirect effects on reading comprehension ability. Our comments on this issue concerning questions and purpose setting are equally relevant here.

Reading Strategies

Whereas several studies have demonstrated that incentives and adjunct questions enhance learning from text, these are conditions imposed on the reader by an outside agent, usually a teacher or a curriculum writer. It would be heartening to discover that children could learn general strategies to apply by themselves thus reducing their dependence on outside agents. Among the attempts to teach students an internalized reading strategy, we include the research on imagery, paraphrasing, self-checking, reading strategy lessons, and corrective feedback.

Pressley (1977) has reviewed research on induced imagery and prose learning. The basic paradigm of this research involves instructing students in how to form mental images of the events specified by the text. The comprehension of students receiving this instruction is then compared to that of students who are not so instructed. There appears to be a clear developmental trend in children's ability to use mental imagery with discourse materials. Children below the fourth grade usually do not benefit from an instruction to form images, either on listening or reading tasks. But in one study Levin and Divine-Hawkins (1974) found that, compared to a no-imagery group, fourth grade students who were told to use an imagery strategy in a *listening* task scored higher on a recall measure. However, no comprehension improvements were noted for students who *read* the same story with imagery instructions.

Whereas younger students have difficulty employing imagery, older students (fifth and sixth graders) have been shown to benefit from imagery instruction in school-like reading tasks (Kulhavy & Swenson, 1975; Rasco, Tenneyson, & Boutwell, 1975). However, younger children (e.g., third and fourth graders) who ordinarily do not benefit from an instruction to image may be taught an imagery strategy. Pressley (1977) gave students 20 minutes of imagery practice with feedback regarding the adequacy of their imaging. In another study, Lesgold, McCormick, and Golinkoff (1975) trained children over several weeks to draw pictures of events described in passages and later to picture the stories in their heads. The results of both studies showed improved comprehension for the imagery-trained subjects. However, in the Lesgold et al. study, students did not spontaneously employ the imagery strategy, but had to be reminded to use it. Moreover, there is evidence from this study and from a study mentioned by Levin (1973) that an imagery strategy will produce improved performance only on certain kinds of reading selections, specifically those containing a large number of concrete references.

These results raise questions about the general utility of imagery strategies. As Pressley (1977) notes, even in the studies which report facilitation from induced imagery, the size of the effects is rather small. Moreover, there appear to be a number of individual difference variables which hinder some students' successful employment of an imagery strategy with prose, including mental retardation (Bender & Levin, 1978), decoding deficiencies (Levin, 1973), and ability to learn from pictures (Levin, Divine-Hawkins, Kerst, & Guttman, 1974).

If learner and text characteristics exert serious limitations on the effectiveness of imagery strategies, are there other means by which readers might successfully create elaborations of text? Doctorow et al. (1978) devised a procedure for text elaboration which would seem not to depend upon the concreteness of text, nor on students' ability to learn from pictures. In their study, sixth grade students were instructed "to generate and to write their own sentence about what happened in the paragraph after they read each paragraph of the story." Even though students received no practice or feedback regarding the adequacy of their sentence paraphrase, they apparently succeeded in performing this task. Compared to a reading-only control group, students in the paraphrase treatment correctly answered 43 percent more questions on a posttest and also completed 43 percent more cloze items one week later. In a replication with lower ability students, the treatment appeared to produce even larger effects with performance improvements of 67 percent. Such promising results from so simple a manipulation call for replication and extension to other reader groups and reading selections. Together with the research on imagery, the Doctorow et al. findings suggest that readers can improve their memory for text by producing mental or verbal elaborations of the events depicted in a passage. How the imagery/sentence generation strategy enhances comprehension or memory for text is not altogether clear. The strategies may be effective because they raise and maintain the reader's attention. Alternatively, they may ensure deeper semantic processing of the information, or may require students to organize and relate the information in the text data to their existing knowledge.

Teaching young readers how and when to systematically employ various reading strategies may affect their success in certain specific reading situations, for example, with highly concrete text. More generally, it would also seem to make them more independent and self-directed learners and more knowledgeable about their own comprehension processes. Brown (in press), describes this phenomenon as the development of metacomprehension. She has identified several skills which mature learners, those who are knowledgeable about their own comprehension and memory abilities, seem to use. In a recent study, Brown, Campione, and Barclay (1978) report an attempt to teach a self-checking skill to mentally retarded adolescents who normally perform poorly on memory and reading tasks. Students learned a rehearsal strategy for memorizing list information. Not only did these students retain this self-checking, rehearsal strategy over a long time, they also spontaneously generalized the strategy to prose materials.

While Brown and Campione focused on what may be seen more as a study skill than a reading skill, the distinction between these areas is sometimes blurred. For instance, are imagery and sentence elaboration strategies reading or study skills? One might argue that in contrast to these reading/study strategies, teaching children a general "reading for meaning" orientation is a purer "reading" strategy. The work of Goodman (1967), a leading spokesman for a meaning emphasis as opposed to a skills and decoding emphasis in the teaching of reading, has prompted several studies on the feedback teachers give to students. In this research, teacher-delivered corrections for oral reading errors either emphasized correct pronunciation and grapheme-phoneme reproduction or emphasized reading for sense. In the former case a teacher might stop a student for every oral reading error, no matter how minor, and require the student to "sound-out" the correction. In the latter case, a teacher might tolerate and even encourage oral reading errors as long as the text modifications made sense and did not alter the author's intended meaning; if teachers did correct errors, they would encourage the student to supply a word which was consistent with the syntax and semantics of the sentence or with the overall passage.

In general, this research suggests that the type of corrections teachers provide can shape students toward either a decoding orientation or a meaning orientation (Hansen, 1977; Piper, 1975; Smith, 1974). However, most of these studies are quasi-experiments and, thus, the strength of their results is somewhat attenuated. In contrast, Niles, Graham, and Winstead (Note 10) recently reported a true experiment in which they contrasted code-oriented corrections of oral reading miscues with a "no-correction" condition. After only four days, fourth grade students appeared to read differently as a function of these feedback variations. On an orally read test passage, students in the no-correction condition made significantly fewer meaning change miscues and also exhibited higher retell scores. Pany, McCoy, and Peters (Note 11) conducted a study of the same instructional conditions as Niles et al., but obtained different results. Remedial reading students assigned to the corrective feedback and no-correction conditions did not differ in answering story-referenced comprehension questions. However, higher functioning remedial students obtained superior retell scores under the no feedback condition.

The findings of Niles et al. are further attenuated by results of an experiment reported by Fleisher (1979). She found no differences on comprehension measures of students taught under the following conditions: 1) corrective feedback for each oral reading error; 2) no corrections during reading, but feedback on answers to comprehension questions; and 3) a combination of corrective feedback on oral reading errors *and* comprehension questions.

Given these conflicting yet intriguing findings, the issue of reading strategies and teacher influence on students' strategy development is an area which deserves far more attention from researchers.

The Current Status of Reading Comprehension Instruction

We have attempted with this review to characterize the research basis for various instructional practices in the teaching of reading comprehension. Given this goal, one might expect us to be in a position to provide some guidance to teachers and curriculum writers about how they should proceed in designing and implementing instruction that will enhance the development of reading comprehension. In our opinion, the current level of knowledge on this topic does not justify many strong prescriptive statements.

A somewhat more optimistic appraisal can be made by focusing on findings from specific studies. While they may not permit our making strong directives to practitioners, certain findings at least provide promising leads for subsequent investigations. In the category of background knowledge variables, the research on pictures (Schallert, in press) may have direct implications for instructional design, and the research on paragraph headings (Doctorow et al., 1978) would appear to warrant more attention from researchers. In the category of linguistic/reasoning variables the research by Dahl (1979) on repeated readings and hypothesis-test training is noteworthy. Weaver's (1979) "chunking" training and the sentence-combining exercises may be effective in developing syntactic knowledge. The task analysis and direct instruction approach represented by the research of Carnine (Note 5) and his colleagues should be of significant interest to teachers and curriculum writers because this approach appears to succeed in teaching the very skills that are often the target of classroom instruction. Finally, within the category of attentional/processing variables, the research on paraphrasing (Doctorow et al., 1978) seems to hold the greatest promise.

On the other hand, the research on cloze practice, advance organizers, purpose setting, speed of single word decoding and imagery suggests that these variables either do not enhance comprehension to any detectable degree, or do so in extremely limited situations (cf. our discussion of imagery).

Stepping back from specific findings to a more general overview, we see that while the literature on comprehension instruction does not answer *the* big question on how to make children better reading comprehenders, it does raise several interesting questions. With the growth in schema-theoretical explanations of comprehension, we need to ask how the notions of world knowledge and schema can be translated into instructionally useful concepts. Recognizing that some first attempts such as advance organizers have not been successful, researchers might rethink the procedures by which educators can identify particular abstract schema or general frameworks which would then be directly and deliberately incorporated into the curriculum. Another relevant area, metacognition, involves strategic behavior for reading situations involving relatively or totally unfamiliar topics, where obvious schemata are absent. Possibly strategies that involve deliberate semantic processing, such as paraphrasing, along with other strategies more directly related to remembering, such as systematic rehearsal, are most appropriate in situations where background knowledge is lacking for a particular topic (Pace, Note 12).

There are other questions which have almost no relationship to theory. Nevertheless, research on these issues might yield some very useful information for practitioners. For example, it would be of interest to examine the effects of the multitude of workbook exercises that are assigned to children in the name of comprehension instruction. Do any or all of these different exercises teach anything and, if so, is that "anything" related to reading comprehension? Some very practical experiments could be designed to address this issue, and their results could have obvious and direct implications for classroom practice.

Another recommendation we make for future instructional research is related to the choice of dependent measures. Many of the studies reviewed above describe their outcome measures in very general terms, for example, as factual questions and inference questions. The failure to employ uniform measures may, in part, account for inconsistent findings across studies. It might be wise to consider also incorporating multiple measures of comprehension into any studies of instructional variables. If multiple measures are not included, the researcher may erroneously conclude either that the instructional treatment did not affect reading comprehension when indeed it did, or that the instructional treatment had widespread impact on reading comprehension, when in fact it did not.

One final comment. We were struck by the increasing volume of research on reading comprehension instruction. However, this research is distinguished more by its breadth (a great diversity of variables are represented) than by its depth. That is, relatively few variables have been studied on more than one occasion by more than one investigator. It is regrettable that some of the most promising variables, in terms of their apparent effects, have been studied on only a single occasion (e.g., Dahl, 1979; Doctorow et al., 1978; Weaver, 1979). The implications of this observation are obvious. Researchers in reading need to establish strong, credible, and generalizable findings if teachers are to be helped in their attempts to foster children's reading development.

References

ANDERSON, R.C. The notion of schemata and the educational enterprise. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge.* Hillsdale, N.J.: Erlbaum, 1977.

AUSUBEL, D.P. *The psychology of meaningful verbal learning.* New York: Grune & Stratton, 1963.

AUSUBEL, D.P. In defense of advance organizers: A reply to the critics. *Review of Educational Research,* 1978, *48,* 251-257.

BALLARD, G.L. The effect of guiding and motivating questions upon the reading comprehension of fourth and fifth grade pupils (Doctoral dissertation, University of Maryland, 1964). *Dissertation Abstracts International,* 1965, *25-08,* 4568. (University Microfilms No. 65-00621.)

BARNES, B.R., & CLAWSON, E.W. Do advance organizers facilitate learning? Recommendations for further research based on an analysis of 32 studies. *Review of Educational Research,* 1975, *45,* 637-659.

BENDER, B.G., & LEVIN, J.R. Pictures, imagery, and retarded children's prose learning. *Journal of Educational Psychology,* 1978, *70*(4), 583-588.

BLUMENFIELD, J., & MILLER, G. Improving reading through teaching grammatical constraints. *Elementary English,* 1966, *43,* 752-755.

BORMUTH, J.R. Readability: A new approach. *Reading Research Quarterly,* 1966, *1,* 79-132.

BORMUTH, J.R., MANNING, J.C., CARR, J.W., & PEARSON, P.D. Children's comprehension of between- and within-sentence syntactic structures. *Journal of Educational Psychology,* 1970, *61,* 349-357.

BRANSFORD, J.D., & JOHNSON, M.K. Contextual prerequisites for understanding: Some investigations of comprehension and recall.

Journal of Verbal Learning and Verbal Behavior, 1972, *11*, 717-726.

BRANSFORD, J.D., & MCCARRELL, N. A sketch of a cognitive approach to comprehension: Some thoughts about understanding what it means to comprehend. In W.B. Weimer & D.S. Palermo (Eds.), *Cognition and the symbolic processes*. Hillsdale, N.J.: Erlbaum, 1975.

BROWN, A. Metacognitive development and reading. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, N.J.: Erlbaum, in press.

BROWN, A.L., CAMPIONE, J.C., & BARCLAY, C.R. *Training self-checking routines for estimating test readiness: Generalization from list learning to prose recall* (Tech. Rep. No. 94). Urbana: Center for the Study of Reading, University of Illinois, 1978.

BROWN, R. *Words and things*. New York: Free Press, 1968.

CHOMSKY, C. *Linguistic development in children from 5 to 10*. Cambridge, Mass.: Radcliffe Research Institute, 1972. (ERIC Document Reproduction Service No. ED 059-196)

CHRISTENSEN, C.M., & STORDAHL, K.E. The effect of organizational aids on comprehension and retention. *Journal of Educational Psychology*, 1955, *46*, 65-74.

COLE, J.N. The effects of non-prose textual characteristics on retention of major concepts and supporting details with and without specific instructions in their use. In P.D. Pearson (Ed.), *Reading: Theory, research and practice* (26th Yearbook of the National Reading Conference), 1977, pp. 232-235.

CRAIK, F.I.M., & LOCKHART, R.S. Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior*, 1972, *11*, 671-684.

CURRIE, W. *An evaluation of two methods of teaching vocabulary through the use of word parts in grades seven, eight, and nine* (Doctoral dissertation, Boston University School of Education, 1963). (University Microfilms No. 643581)

DAHL, P.R. An experimental program for teaching high speed word recognition and comprehension skills. In J.E. Button, T.C. Lovitt, & T.D. Rowland (Eds.), *Communications research in learning disabilities and mental retardation*. Baltimore: University Press, 1979.

DALE, E., RAZIK, T., & PETTY, W. *Bibliography of vocabulary studies*. Columbus: Ohio State University, 1973.

DAUGHERTY, J. The effects of the cloze procedure and interspersed questions as an aid to reading comprehension (Doctoral dissertation, University of Akron, 1971). *Dissertation Abstracts International*, 1971, 32-02A 664. (University Microfilms No. 71-20275)

DOCTOROW, M., WITTROCK, M.C., & MARKS, C. Generative processes in reading comprehension. *Journal of Educational Psychology*, 1978, *70*, 109-118.

DURKIN, D. *What classroom observations reveal about reading comprehension instruction* (Tech. Rep. No. 106). Urbana-Champaign: Center for the Study of Reading, University of Illinois, 1978.

FINCKE, W. The effect of asking questions to develop purposes for reading on the attainment of higher levels of comprehension in a population of third grade children (Doctoral dissertation, Temple University, 1968). *Dissertation Abstracts International*, 1968, 29/06A,1778. (University Microfilms No. 68-14131)

FISHER, K.D. *An investigation to determine if selected exercises in sentence-combining can improve reading and writing* (Doctoral dissertation, Indiana University, 1973). (University Microfilms No. 74-2586)

FLEISHER, L.S. *The effect of word-emphasis and comprehension-emphasis instruction on reading performance.* Unpublished doctoral dissertation, University of Illinois, 1979.

FLEISHER, L.S., JENKINS, J.R., & PANY, D. Effects on poor readers' comprehension of training in rapid decoding. *Reading Research Quarterly*, 1979, *15*, 30-48.

FRIES, C.C. *Linguistics and reading.* New York: Holt, Rinehart & Winston, 1963.

GOODMAN, K. Reading: A psycholinguistic guessing game. *Journal of the Reading Specialist*, 1967, *1*, 126-135.

GOODMAN, Y.M., & BURKE, C.L. *Reading miscue inventory manual: Procedure for diagnosis and evaluation.* London: Macmillan, 1972.

GUICE, B. The use of cloze procedures for improving comprehension of college students. *Journal of Reading Behavior*, 1969, *1*, 81-92.

HAFNER, L.E. A one month experiment in teaching context aids in fifth grade. *The Journal of Educational Research*, 1965, *58*, 472-474.

HANSEN, C. *The generalization of skills and drills vs. corrective feedback instruction to the independent reading performance of intermediate aged learning disabled boys* (Doctoral dissertation, University of Washington, 1977). (University Microfilms No. 77-582)

HANSEN, C., & LOVITT, T. The relationship between question type and mode of reading on the ability to comprehend. *Journal of Special Education*, 1976, *10*, 53-60.

HANSEN, C., & LOVITT, T. An applied behavior analysis approach to reading comprehension. In J.T. Guthrie (Ed.), *Cognition, curriculum, and comprehension.* Newark, Del.: International Reading Association, 1977.

HARRIS, A. *How to increase reading ability* (5th ed.). New York: David McKay, 1970.

HEITZMAN, A., & BLOOMER, R. The effects of non overt reinforced cloze procedures upon reading comprehension. *Journal of Reading*, 1967, *11*, 213-223.

HUGHES, T.O. *Sentence combining: A means of increasing reading comprehension*. Bloomington, Ind.: Eric Clearinghouse on Reading, 1975. (Eric Document Reproduction Service No. ED 112421)

JACKSON, J., & DIZNEY, H. Intensive vocabulary training. *Journal of Developmental Reading*, 1963, *6*, 221-229.

JENKINS, J.R., BARKSDALE, A., & CLINTON, L. Improving reading comprehension and oral reading: Generalization across behaviors, settings, and time. *Journal of Learning Disabilities*, 1978, *11*(10), 607-617.

JENKINS, J.R., & PANY, D. Teaching reading comprehension in the middle grades. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension*. Hillsdale, N.J.: Erlbaum, in press.

JENKINS, J.R., PANY, D., & SCHRECK, J. *Vocabulary and reading comprehension* (Tech. Rep. No. 100). Champaign-Urbana: Center for the Study of Reading, University of Illinois, 1978.

KATZENMEYER, C., & VAN BLARICOM, G. *Toward the development of a set of measures of deductive reasoning in the elementary school* (Working Paper No. 172, Wisconsin Research and Development Center for Cognitive Learning and Instruction). Madison: University of Wisconsin, 1976.

KENNEDY, D., & WEENER, P. Visual and auditory training with cloze procedures to improve reading and listening comprehension. *Reading Research Quarterly*, 1973, *8*, 824-841.

KOLERS, P.A. Some model of representation. In P. Pliner, L. Kramer, & T. Alloway (Eds.), *Communication and affect: Language and thought*. New York: Academic Press, 1973.

KULHAVY, R.W., & SWENSON, I. Imagery instructions and the comprehension of text. *British Journal of Educational Psychology*, 1975, *45*, 47-51.

LABERGE, D., & SAMUELS, S.J. Toward a theory of automatic information processing in reading. *Cognitive Psychology*, 1974, *6*, 293-323.

LAHEY, B., MCNEES, M., & BROWN, C. Modification of deficits in reading for comprehension. *Journal of Applied Behavior Analysis*, 1973, *6*, 475-480.

LANDRY, D.L. The effect of organizational aids on reading comprehension (Doctoral dissertation, University of Connecticut, 1966). *Dissertation Abstracts International*, 1967, 27/10A, 3229.

(University Microfilms No. 67-03898)

LESGOLD, A.M. Variability in children's comprehension of syntactic structures. *Journal of Educational Psychology*, 1974, *66*(3), 333-338.

LESGOLD, A.M., MCCORMICK, C., & GOLINKOFF, R.M. Imagery training and children's prose learning. *Journal of Educational Psychology*, 1975, *67*, 663-667.

LEVIN, J.R. Inducing comprehension in poor readers: A test of a recent model. *Journal of Educational Psychology*, 1973, *65*, 19-24.

LEVIN, J.R., & DIVINE-HAWKINS, P. Visual imagery as a prose learning process. *Journal of Reading Behavior*, 1974, *6*, 23-30.

LEVIN, J.R., DIVINE-HAWKINS, P., KERST, S.M., & GUTTMANN, J. Individual differences in learning from pictures and words: The development and application of an instrument. *Journal of Educational Psychology*, 1974, *66*, 296-303.

LEVIN, J.R., & LESGOLD, A.M. On pictures in prose. *Educational Communication and Technology*, in press.

LIEBERMAN, J. *The effect of direct instruction in vocabulary concepts on reading achievement.* Bloomington, Ind.: ERIC Clearinghouse on Reading, 1967. (ERIC Document Reproduction No. ED 010985)

LOUTHAN, V. Some systematic grammatical deletions and their effects on reading comprehension. *English Journal*, 1965, *54*, 295-299.

LOVITT, T., & HANSEN, C. The use of contingent skipping and drilling to improve oral reading and comprehension. *Journal of Learning Disabilities*, 1976, *9*, 481-487.

LOWERRE, G., & SCANDURA, J. Conceptually based development and evaluation of individualized materials for critical reading based on logical inference. *Reading Research Quarterly*, 1973-1974, *9*, 187-205.

MEYER, B.J.F. *The organization of prose and its effect on recall.* Amsterdam: North-Holland Publishers, 1975.

NARDELLI, R. Some aspects of creative reading. *Journal of Educational Research*, 1957, *50*, 495-508.

OTTERMAN, L.M. The value of teaching prefixes and word-roots. *Journal of Educational Research*, 1955, *48*, 611-616.

PANY, D. *The effects of vocabulary instruction on knowledge of word meanings and reading comprehension of remedial readers.* Unpublished doctoral dissertation, University of Illinois, 1978.

PANY, D., & JENKINS, J.R. Learning word meanings: A comparison of instructional procedures and effects on measures of reading comprehension with learning disabled students. *Learning Disability Quarterly*, 1978, *1*, 21-32.

PEARSON, P.D. The effects of grammatical complexity on children's comprehension, recall and conception of certain semantic relations. *Reading Research Quarterly*, 1974-1975, *10*(2), 155-192.

PEARSON, P.D., HANSEN, J., & GORDON, C. *The effect of background knowledge on young children's comprehension of explicit and implicit information* (Tech. Rep. No. 116). Champaign-Urbana: Center for the Study of Reading, University of Illinois, 1979.

PERFETTI, C.A. *Language comprehension and fast decoding: Some psycholinguistic prerequisites for skilled reading comprehension.* In J.T. Guthrie (Ed.), Cognition, curriculum, and comprehension. Newark, Del.: International Reading Association, 1977.

PERFETTI, C.A., & HOGABOAM, T. The relationship between single word decoding and reading comprehension skill. *Journal of Educational Psychology*, 1975, *67*, 461-469.

PETTIT, N.T. Effects of reading for a given purpose on literal and inferential comprehension (Doctoral dissertation, University of Missouri, 1970). *Dissertation Abstracts International*, 1971, 31/10A, 5255. (University Microfilms No. 71-08376)

PIPER, C.J. A measure of the effectiveness of instructional strategies based on psycholinguistic principles of teaching reading when applied to a program to aid seventh and eighth grade underachievers (Doctoral dissertation, Wayne State University, 1975). *Dissertation Abstracts International*, 1975, *36*, 2602A. (University Microfilms No. 75-25-296)

PRESSLEY, M. Imagery and children's learning: Putting the picture in developmental perspective. *Review of Educational Research*, 1977, *47*(4), 585-622.

RASCO, R.W., TENNYSON, R.D., & BOUTWELL, R.C. Imagery instructions and drawings in learning prose. *Journal of Educational Psychology*, 1975, *67*, 188-192.

RICHMOND, M.G. The relationship of the uniqueness of prose passage to the effect of question placement and question relevance on the acquisition and retention of information. In W.D. Miller & G.H. McNinch (Eds.), *Reflection and investigations of reading* (25th Yearbook of the National Reading Conference), 1976, pp. 268-279.

ROBINSON, H.M. Developing critical readers. In R.G. Stauffer (Ed.), *Dimensions of critical reading,* Proceedings of the Annual Education and Reading Conferences, 1964, *11*, 1-12.

ROYER, J.M., & CABLE, G.W. Illustrations, analogies and facilitative transfer in prose learning. *Journal of Educational Psychology*, 1976, *68*, 205-209.

SAMUELS, S.J. Effects of pictures on learning to read, comprehension and attitudes. *Review of Educational Research*, 1970, *40*, 397-407.

SAMUELS, S.J., DAHL, P., & ARCHWAMETY, T. Effect of hypothesis/test training on reading skill. *Journal of Educational Psychology*, 1974, *66(6)*, 834-844.

SCHALLERT, D.L. The role of illustrations in reading comprehension. In R.J. Spiro, B.C. Bruce, & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension.* Hillsdale, N.J.: Erlbaum, in press.

SCHNEYER, J.W. Use of the cloze procedure for improving reading comprehension. *The Reading Teacher*, 1965, *19*, 174-179.

SMILEY, S.S., OAKLEY, D.D., WORTHEN, D., CAMPIONE, J.C., & BROWN, A.L. Recall of thematically relevant material by adolescent good and poor readers as a function of written versus oral presentation. *Journal of Educational Psychology*, 1977, *69*(4), 381-387.

SMITH, B.V. A descriptive study of the effectiveness of reading strategy lessons as indicated by the reading miscue inventory profiles of selected below average readers (Doctoral dissertation, Virginia Polytechnic Institute and State University, 1974). *Dissertation Abstracts International*, 1974, *35*, [illegible]03A. (University Microfilms No. 75-3468)

SMITH, F. *Psycholinguistics and reading.* New York: Holt, Rinehart & Winston, 1973.

SNAVELY, A.E. The effectiveness of purpose statements and marginal notes as aids to reading comprehension (Doctoral dissertation, State University of Iowa, 1961). *Dissertation Abstracts International*, 1962, 22-08A, 2711. (University Microfilms No. 61-6511)

SPACHE, G.D. *Toward better reading.* Champaign, Ill.: Garrard Publishing, 1966.

SPIRO, R. *The etiology of reading comprehension styles* (Tech. Rep. No. 124). Champaign-Urbana: Center for the Study of Reading, University of Illinois, 1978.

STICHT, T., BECK, L., HAUKE, R., KLEIMAN, G., & JAMES, J. *Auding and reading: A developmental model.* Alexandria, Va.: Human Resources Research Organization, 1974.

SWENSON, I., & KULHAVY, R.W. Adjunct questions and the comprehension of prose by children. *Journal of Educational Psychology*, 1974, *66*, 212-215.

THORNDYKE, P. Cognitive structures in comprehension and memory of narrative discourse. *Cognitive Psychology*, 1977, *9*, 77-110.

WEAVER, P.A. Improving reading comprehension: Effects of sentence organization instruction. *Reading Research Quarterly*, 1979, *15*(1), 129-146.

WEINTRAUB, S. The question as an aid in reading. *The Reading Teacher*, 1969, *22*, 751.

WOLF, W., KING, M., & HUCK, C. Teaching critical reading to elementary school children. *Reading Research Quarterly*, 1968, *3*(4), 435-498.

YOST, M., AVILA, L., & VEXLER, E.B. Effects on learning of post instructional responses to questions of varying degrees of complexity. *Journal of Educational Psychology*, 1977, 69, 399-408.

Reference Notes

1. Mason, J., & Boggs, J. *A survey of classroom reading instruction—structure and activities in grades two through five.* Manuscript in preparation, Center for the Study of Reading, University of Illinois, Urbana, 1978.
2. Clements, J., Stevens, C.W., Kameenui, E.J., & Carnine, D.W. *Instructional procedures for identifying and interpreting characters' motives during narrative reading.* Unpublished manuscript, University of Oregon, 1979.
3. Kameenui, E.J., & Carnine, D.W. *Vocabulary variation in passage reading.* Unpublished manuscript, University of Oregon, 1979.
4. Sindelar, P. *A comparison of three methods of supplemental reading instruction.* Unpublished manuscript, Department of Special Education, Pennsylvania State University, 1978.
5. Carnine, D.W., Prill, N., & Armstrong, J. Teaching slower performing students general case strategies for solving comprehension items. In *Technical Report 1978-2, Formative research on direct instruction.* Eugene: University of Oregon, 1978.
6. Coyle, G., Kameenui, E.J., & Carnine, D.W. *A direct instruction approach to teaching the utilization of contextual information in determining the meaning of an unknown word embedded in a passage.* Unpublished manuscript, University of Oregon, undated.
7. Kameenui, E.J., Carnine, D.W., & Maggs, A. *Task analysis of instructional procedure for reversible passive and clause constructions.* Unpublished manuscript, University of Oregon, undated.
8. Woolfson, N., Kameenui, E.J., & Carnine, D.W. *Direct instruction procedures for making inferences with variations in the explicitness, complexity, and dispersal of information—An experimental study.* Unpublished manuscript, University of Oregon, 1979.
9. Becker, J., & Lilly, M.S. *The relationship between listening and reading for junior high students.* Unpublished manuscript, Department of Special Education, University of Illinois at Urbana-Champaign, 1979.
10. Niles, J.A., Graham, R.T., & Winstead, J.C. *The effects of teacher feedback responses on children's oral reading performance.* Paper presented at the International Reading Association Convention, Miami, 1977.

11. Pany, D., McCoy, K.M., & Peters, E. *Teacher feedback vs. uninterrupted reading: Effects on reading comprehension* (Working Paper No. 1). Tempe: Arizona State University, 1979.
12. Pace, A.J. *The influence of world knowledge on children's comprehension of short narrative passages.* Paper presented at the annual meeting of the American Educational Research Association, Toronto, March 1978.

Academic Learning Time and Reading Achievement[1]

David C. Berliner
University of Arizona

If one is willing to accept the standardized reading achievement test score as a criterion for effective instruction, then there is now sufficient evidence to say that some of the variables associated with successful instruction in reading at the elementary grades are known. To interpret this evidence requires familiarity with just a handful of concepts: *allocated time, engaged time, success rate, academic learning time, opportunity to learn, content coverage, curriculum-test congruence*, and *direct instruction*. These concepts, as they apply to elementary grade reading instruction, are described in more detail below.

Allocated Time

Table 1 presents data describing time allocations in reading and language arts activities from 25 second grade and 21 fifth grade classes. (The complete study, including descriptions of the sample, instrumentation, and data analysis, is reported in Fisher, Filby, Marliave, Cahen, Dishaw, Moore, & Berliner, 1978b.) These data are based on records of classroom activity kept for most of the school days from January to May of a recent school year in a sample of schools from the San Francisco,

---

[1] This paper is based, in part on the results of a multi-year study called the Beginning Teacher Evaluation Study. The study and papers issued from the project were a joint effort of David C. Berliner, Charles W. Fisher, Nikola N. Filby, Richard Marliave, Leonard S. Cahen, Marilyn Dishaw, and Jeffry E. Moore.

California, Bay area (Fisher, Filby, & Marliave, 1977). What is immediately obvious from these data is the enormous variability of allocated time for reading and reading related activities. In the second grade, in classroom 4, students are provided with a mean of 47 minutes a day for reading. In contrast, similar students in classrooms 11 and 17 each received an average daily allocation of 118 minutes for reading. The same phenomenon is found in the fifth grade data. Students in classroom 6 received a little over an hour a day for reading, while in classroom 10 and classroom 12 the average student is exposed to 2 hours and 17 minutes of reading instruction and related reading activities. These very large differences in the duration of reading instruction in different classes are great enough to affect scores on standardized reading tests (Fisher et al., 1978b). It is difficult to determine what the "proper" or "upper" limits of allocated time for reading should be in the elementary grades. But it is *not* difficult to believe, from these data, that some teachers have allocated too little time for reading instruction.

Another item of interest in the data of Table 1 is the standard deviation of daily allocated time in reading. In the fifth grade classes, class 8 and class 18 had almost identical means for allocated reading time, but they show quite different standard deviations. The managerial behavior of the teacher in classroom 18 was such that very little variability from a set routine is shown. In class 8, much more variability in the daily pattern of classroom organization is shown. Such differences in style of classroom management also affect achievement. The large between-class differences in standard deviations and total allocated time sometimes occur because transition time (the time students spend finishing one activity, moving, and getting ready for another activity), wait time (the time students spend waiting for help or directions), and time spent in behavioral management in some classes is very high or variable from day to day. These data indicate that the time allocated for academic instruction in a school day can easily slip away when a teacher cannot keep the transitional time, wait time, and behavioral problems to a minimum. Any sensible manager knows that. Somehow,

Table 1
Mean Time and Standard Deviation (in Minutes) Devoted to Reading and Language Arts Activities by Class and by Grade

Second Grade		Fifth Grade	
Class	Time	Class	Time
1	106 (3)	1	102 (3)
2	79 (8)	3	80 (2)
3	83 (4)	4	92 (4)
4	47 (8)	5	110 (8)
5	96 (5)	6	68 (7)
6	65 (5)	8	104 (13)
7	93 (7)	9	127 (4)
8	75 (6)	10	137 (11)
9	90 (1)	11	119 (5)
10	93 (6)	12	137 (9)
11	118 (7)	14	130 (4)
12	83 (5)	16	95 (8)
13	87 (1)	17	108 (15)
14	103 (6)	18	106 (2)
15	96 (3)	19	108 (6)
16	66 (3)	21	129 (9)
17	118 (3)	23	121 (6)
18	101 (4)	24	95 (10)
19	90 (2)	25	129 (3)
20	80 (4)	26	102 (8)
21	82 (4)	27	88 (3)
22	86 (6)	——	——
23	85 (6)	——	——
24	90 (2)	——	——
25	78 (5)	——	——
Total Sample	88 (16)	Total Sample	74 (11)

Source: Fisher, Filby, & Marliave, 1977.

however, in many classes, there is a lack of attention to classroom management that results in considerable inefficiency, and reduced achievement on standardized tests of reading.

When studying instructional time it is also important to know how much time is allocated to particular areas of the reading curriculum. Table 2 presents such data for four fifth grade classes. These data were collected from teacher logs kept on a sub-sample of students in the class, over about 90 days of instruction. Considerable variability in the allocations of time to particular content areas of fifth grade reading is shown. (Similar results were found at other grade levels and for other subject matter.)

Different philosophies of education result in different beliefs about what is important for students to learn. These beliefs, along with the teacher's likes and dislikes for teaching certain areas, result in some interesting differences in the *functional* curriculum of a class. For example, from Table 2 it can be seen that classroom C spends dramatically more time on comprehension in reading than any of the other three fifth grade classes. In classroom D silent reading and spelling were emphasized, as judged from the dramatically greater allocation of time to those content areas, in contrast to the average amount of time each student of classes A, B, and C received. And oral reading hardly seemed to be of interest to the teacher of class B, at least that is what can be concluded when the data from class B are compared with the data from the other fifth grade classes.

These rather significant differences in the functional classroom curriculum do result in considerable differences in achievement (Fisher et al., 1978b). If students in these fifth grade classes were part of some end-of-year statewide testing program, where drawing inferences from paragraphs of prose was tested, as it often is, one might well expect that students in classroom C would show superior performance when contrasted to similar students in the other fifth grade classes.

This brief examination of selected data presenting estimates of classroom allocated time shows clearly that some teachers spend considerably more time instructing in particular reading content areas than other teachers, and some teachers allocate considerably more total instructional time to reading than do other teachers. These differences, put into experimental terminology, represent clear differences in the type and in the duration of treatment. And we should expect that when type of treatment and duration of treatment are varied, achievement will vary. Our data confirm this. Other things being equal, the more time allocated to a content area of reading, the higher the academic achievement in that content area.

Table 2
Allocated Pupil Time (in Minutes) in Content Areas of Reading for Four Fifth-Grade Classes

	Curriculum Content Areas	Classes A	 B	 C	 D
Word structure	Root words and affixes	250	112	126	103
	Syllables	67	60	102	212
Word meaning	Synonyms	95	152	10	119
	Pronoun reference	0	0	9	56
	Other word meaning	558	949	1042	615
Comprehension	Verbatim (no rephrasing)	206	329	188	325
	Translation (paraphrase)	122	151	1649	383
	Inference/Synthesis	235	252	1432	306
	Identifying main items	153	243	943	326
	Evaluation of fact and opinion	5	0	66	56
	Other comprehension	196	325	1368	239
Reading practice	Oral reading	604	63	885	305
	Silent reading	1083	724	956	3640
	Reading in content areas	505	256	400	284
Related reading activities	Spelling	694	847	664	1415
	Grammar	242	183	859	413
	Creative writing	56	343	98	573
	Study skills	472	669	270	171
	Other	207	687	1317	426

Sources: Dishaw, 1977a; Dishaw, 1977b; Filby & Marliave, 1977.

Engaged Time

Table 3 presents data on the average percentage of time students are engaged during reading instruction in the four fifth grade classes described in Table 2. These data are from observer records and not from teacher logs. Previous work revealed that teachers can keep accurate records of allocated time, but that classroom observers are necessary to obtain accurate records of engaged time (Marliave, Fisher, Filby, 1977). In examining these data it appears that the percentage of time students are engaged is relatively high. This is an artifact of the observational system that was in use. The observation system required that transition time and certain other classroom phenomena be coded as separate events. Thus, the data on engagement rates are for the time spent in reading, after a class has settled down and before the class starts to put their work away. If engagement were coded for the entire time block denoted by teachers as reading time, the engaged time rates would be considerably lower because during transitions or when waiting for help students are usually not engaged. Still, variability between classes is noted for this important variable. The engagement rates in these four fifth-grade classes vary from 75 percent to 84 percent during reading instruction. This range was much larger in the total sample of classes studied.

The average number of minutes per day allocated for instruction, multiplied by the engagement rate, provides liberal estimates of the number of engaged minutes per day, per student. These data are found in Table 3. The range in these four classes is between 48 and 119 minutes per day of engaged time. These are dramatic differences, *differences of 100 percent or more*, in the engaged time students allot to learn their reading. And these differences in engagement have been consistently related to differences in achievement. There is nothing very startling here. If students do not pay attention, they do not learn much.

In most districts we may assume that a school year is about 180 days. This figure must be reduced by absences of teachers and students, strikes, bussing difficulties, the difficulties of instruction before Christmas and Easter breaks, the testing at the beginning and end of the school year, and other factors. A reasonable, perhaps even a liberal, estimate of the "functional" school year may be about 150 days. Accumulating the engaged minutes per day over these 150 days gives an estimate of the engaged instructional time allotted by students to the academic curriculum during the entire school year. Table 3 also presents these data. In the four fifth grade classes, with reasonably mature and independent learners, between 120 and 298 cumulative hours per school year are noted for *all* areas of reading. It is worth noting that these teachers were volunteers who were open enough about their teaching to permit regular and extensive observation of their classes. Thus, we estimate that the data on engaged time are markedly higher in the sample than in

Table 3
Engaged Time, Percent of Time Students Are Working with Easy Material, and Academic Learning Time for Four Fifth-Grade Reading Classes

	Classes			
	A	B	C	D
Percent of time students engaged	82	77	84	75
Engaged minutes per day	48	51	119	101
Percent of time students are in material of easy difficulty level	51	61	47	58
Academic learning time per day in minutes	24	31	56	59
Engaged hours per 150 day school year	120	128	298	283
Academic learning time, in hours, per 150 day school year	60	78	140	148

the population of interest. We have reason to suspect that in *many* fifth grade classes cumulative engaged time in reading is well under 100 hours for the entire school year. With younger children, say second graders, the total time students are engaged in the reading curriculum for the entire school year is considerably lower, and just as variable. We have reason to suspect that in *many* second grade classes cumulative engaged time in reading is well under 70 hours for the entire school year.

As these data come to light some important questions must be asked. For example, what should be expected in the way of engaged time for 30 students and one teacher, working together throughout the school year? What are the expectations for instructional time held by parents and school board members as they make policy to educate the young of a community? Because these new estimates of classroom allocated and engaged time do not conform to the prevailing beliefs that exist among people who manage and support education, either those beliefs must be changed or instructional practices must be altered.

Success Rate

Three rather broad categories were used in the Beginning Teacher Evaluation Study to define the difficulty level of the material or activities that were worked on by students (Fisher et al., 1978b). In "high success," the student understands the task and makes only occasional careless errors. In "medium success," the student has partial understanding but makes some substantive errors. In "low success," the student does not understand the task at all. These categories coincide with common sense notions of "easy," "medium," and "hard."

High success rate. The findings consistently point out the positive effects of school tasks yielding high success rates (easy materials, providing a low error rate). Other research on instructional design has stressed the importance of high success rates. High success rate in scholastic activities has also been found to be one factor that contributes to high levels of student self-esteem.

The average student in the Beginning Teacher Evaluation Study spent about half the time working on tasks that provided high success. Students who spent more time than the average in high success activities had higher achievement scores in the spring, better retention of learning over the summer, and more positive attitudes toward school. From these data, one might recommend that students spend somewhat more than half their time on tasks they can carry out with high success. Sixty or seventy percent might be reasonable.

The idea of success rate may be more understandable if one thinks about the cyclical nature of learning. Learning is a process of moving from not knowing to knowing. Most likely, when new material is introduced the student will not understand completely and will make some errors. Guided practice and/or explanation help the student understand, and she or he comes to make fewer errors. Eventually, the student will perform correctly, although probably with some effort. Learning will become well established and further work will be practice or review. This stage could be viewed as one of consolidation. At some later point, the student knows the material so well that further practice is of minimal value and it is time to move on to something new. The results of the Beginning Teacher Evaluation Study suggest that for learning of basic skills in the elementary grades, the stage of *successful practice* (consolidation) is particularly important, so that concepts and procedures are thoroughly mastered. Apparently some teachers do not devote sufficient time to this stage.

While these data lead one to emphasize the importance of giving students ample opportunity for successful practice, one must also point out that it would not be desirable for students to spend all of their time on tasks they can perform completely correctly. Common sense suggests that too high a rate of "high success" work would be deleterious (boring, repetitive, time wasting, etc.). Probably, some balance between "high success" and more challenging work is appropriate. Also, it was found that older students and/or students who were generally skilled at school learning benefited from a smaller percentage of time at the high success level. Apparently these students had learned to problem solve, to take a task they did not completely understand and work it out. These students may enjoy the challenge of more difficult materials, as long as they eventually experience success.

Low success rate. When students worked with materials or activities that were categorized as "hard," yielding a low success rate, achievement was lower. In the Beginning Teacher Evaluation Study, no teacher assigned a high proportion of materials that were exceptionally hard for students. However, some students worked on materials judged to be excessively difficult for them as much as 20 percent of the time. Other students never worked at a low success rate. Students who were observed to spend more time on excessively difficult materials generally learned less than other students. It is seldom, if ever, desirable for students to be given tasks where they experience low success.

Academic Learning Time

Academic Learning Time (ALT) was the research variable of most interest in the Beginning Teacher Evaluation Study. This variable is defined as the time a student is engaged with academic materials or activities that yield a high success rate. One component of ALT is the engagement rate of students. Another is the level of difficulty of the material that is attended to by a student. Theoretical and empirical evidence suggests that classroom learning occurs primarily with materials that are of an easy level of difficulty. Materials that are too hard for a student do not add much to his or her acquisition of the concepts, skills, and operations that are required of students in a particular grade level. Nor do they allow for practice, repetition, and overlearning. These are important concerns if retention is to be maximized. Table 3 presents information on the percentage of time that students are working with relatively easy material. These data are ratings made by observers in classrooms. As shown in Table 3, for fifth-grade reading the range is between 47 percent and 61 percent. Multiplying the engaged minutes per day by the percent of time students are assigned work that yields low error rates provides an estimate of ALT per day. These data are also provided in Table 3.

As noted above, the typical academic school year of 180 days may be considered to be a functional school year of 150 days. The last line in Table 3 presents academic learning time, in hours, for a school year of 150 days. In fifth grade reading the range is from 60 hours per school year to 148 hours per school year. In these four classes, differences of many hundreds of percent in accumulated ALT are noted. In the total sample studied, the range of ALT is considerably larger. It should again be noted that all the elementary school teachers in this sample were volunteers. These data, if they could be obtained from a nonvolunteer sample, would most likely show even more between class variability.

If academic learning time is a major factor in acquiring the knowledge and skill required to master the curriculum of a particular grade level, for a particular content area, one can see that the school year does not contain as much ALT as might be desired. If our concerns about instruction are correct, there are many, many classes where there is not sufficient time for students to master the curriculum that has been chosen for them. The implications of this situation for learning are quite important and are discussed next.

Academic Learning Time and Achievement

A major finding of the Beginning Teacher Evaluation Study is that increases in Academic Learning Time are associated with increases in student achievement. The practical importance of Academic Learning Time in relationship to achievement is illustrated in Table 4, using examples from the analysis of grade two reading instruction (Fisher, Berliner, Filby, Marliave, Cahen, Dishaw, & Moore, 1978a). This table displays total reading scores in October, student engaged time with a high success rate (Academic Learning Time) in reading over the period from October to December, and estimated total reading scores in December. The reading scores in December are estimated from linear regressions that use the test scores in October and the intervening Academic Learning Time to estimate the expected December test scores. Total reading scores in October and December are shown both in terms of raw scores out of 100 items (percent of items correct) and in terms of percentile rank among students in this study. Student engaged time with a high success rate (Academic Learning Time) in reading is shown both in terms of total time over the five week inter-test period and in terms of the corresponding average daily time (in minutes).

Reading across the rows in this table from left to right, one can see that given a particular reading score in October (two left columns) and a particular amount of Academic Learning Time (middle two columns), what the estimated reading score in December is (two right columns). It should be recognized, however, that students who started with a given score in October and experienced a given amount of Academic Learning Time during the intervening period did not always attain the same score in December. Therefore, it was necessary, for the puposes of this table, to arrive at some estimate of a "typical" or expected score in December, given certain initial scores and some amount of Academic Learning Time. In Table 4, the estimated December raw scores have a standard error of 11.4. This indicates that two-thirds of the pupils with an estimated score of, say, 40 will have scores between 28.6 and 51.4.

Table 4 shows that substantial increases in Academic Learning Time are associated with important increases in achievement. Consider the student who started the period with a grade two reading score that was average (50th percentile, see the top three rows). If this student experiences the average amount of Academic Learning Time (573 minutes total, or 23 minutes per

Table 4
Academic Learning Time and Student Achievement: Examples from Grade Two Reading Based on the Beginning Teacher Evaluation Study, Phase III-B Results

Reading Score at First Testing (October)		Student Engaged Time in Reading with High Success Rate		Estimated Reading Score, Second Testing (December)	
Raw Score (out of 100)	Percentile	Total Time Over 5 Weeks (Minutes)	Average Daily Time (Minutes)	Raw Score (out of 100)	Percentile
36	50	100	4	37	39
36	50	573	23	43	50
36	50	1300	52	52	66
16	17	100	4	20	15
16	17	573	23	25	21
16	17	1300	52	35	36

Notes:
1. An average of 25 school days occurred between the first and the second testing.
2. The B reading scores are estimated via linear regression.
3. The values of all variables in this table are within the ranges actually obtained in the sample.
4. The average engaged time with high success rate in grade two reading for the inter-test period was 573 minutes.

day in reading), then the student can be expected to show average reading achievement in December (50th percentile again). Note also that the "average" student with "average" Academic Learning Time does show considerable learning in terms of predicted raw scores.

Table 4 also indicates that if this average student (in terms of October's test score) experienced only four minutes per day of Academic Learning Time (100 minutes total for the inter-test period), then she or he would be expected to show *almost no* change in raw scores (36 out of 100 correct in October, 37 out of 100 in December) and would decline considerably in relative terms (50th percentile in October, 39th percentile in December). However, if the same student experienced very large amounts of Academic Learning Time, 52 minutes per day in this example, then he or she could be expected to answer almost 50 percent more items correctly on the December test than on the October test (36 out of 100 correct in October, 52 out of 100 correct in December). Furthermore, in this situation the student would show considerable improvement in reading achievement relative to the other students in the study (50th percentile in October, 66th percentile in December). Thus, the student with large amounts of Academic Learning Time benefits substantially.

It may appear that this range from 4 to 52 minutes per day is unrealistically large. *However, these values actually occurred in the classes in the study.* Furthermore, one can easily imagine how either 4 or 52 minutes per day of Academic Learning Time might come about. If 50 minutes of reading instruction per day is allocated to a student who pays attention only about a third of the time and only one-fourth of the student's reading time is at a high level of success, then the student will experience only about 4 minutes of engaged reading at a high success level. Similarly, if 100 minutes per day are allocated to reading for a student who pays attention 85 percent of the time, at a high level of success for almost two-thirds of that time, then she or he will experience about 52 minutes of Academic Learning Time per day.

In summary, large differences in Academic Learning Time are associated with very important changes in predicted achievement levels. These large differences in Academic Learning Time are well within the range that was actually observed for the Beginning Teacher Evaluation Study. In addition, one can easily imagine how these differences in Academic Learning Time could occur in *realistic* situations. Therefore, Academic Learning Time is shown to be of considerable *practical* importance in terms of its relationship to achievement.

Academic Learning Time and Attitude

The data from the Beginning Teacher Evaluation Study revealed that students with high and low rates of allocated and engaged time were equally likely to have positive or negative attitudes toward the subject matter and the school. Educators are naturally concerned about whether greater than average time in academic pursuits or greater than average rates of attending will result in negative attitudes. In the Beginning Teacher Evaluation Study, that did not happen. There was one consistent, positive trend in the data. It appears that students experiencing high rates of success are somewhat more likely to have positive attitude toward reading and school.

Opportunity to Learn, Content Coverage, and Curriculum-Test Congruence

In the recently completed Instructional Dimensions Study (Cooley & Leinhardt, 1978), a very well-done search for effective classroom processes, the variable "opportunity to learn" was extensively examined. Opportunity to learn was a composite variable. A classroom would score high on the variable if it had: lower enrollment, higher rates of attendance, higher allocated times in reading and mathematics, fewer transfers in or out, and higher rates of on-task behavior. Thus, in this study, the opportunity variable overlapped with the allocated and engaged time measures mentioned above.

The variable of opportunity to learn also was defined in terms of curriculum overlap—an estimate of the overlap between what was taught and what was in the end-of-the-year achievement test. This is the issue of the degree of congruence between the curriculum that is taught and the achievement test used to measure mastery of the curriculum. Another and related aspect of opportunity to learn, not studied directly in Cooley and Leinhardt, is content coverage. However, others have addressed the content coverage issue directly (Borg, 1978; McDonald & Elias, 1976).

The results of these studies are now quite clear and consistent (cf. Rosenshine & Berliner, 1978). *Opportunity to learn, content coverage, and curriculum-test congruence are important variables, discriminating between more and less effective teachers.*

Cooley and Leinhardt say it this way: "In summary, the major generalization with respect to classroom processes must be that the most useful construct in explaining achievement gain is the opportunity that the children had to learn the skills assessed in the achievement test" (p. 32).

Opportunity is high when allocated time in some content areas is high, engaged time in that content area is high, content coverage in that curricular area is broad, and that the time and content choices match the depth and breadth of the achievement tests used in assessing the instruction.

Direct Instruction

All these trends in the data can be brought together under the still nebulous but semantically rich concept called "direct instruction" (Berliner & Rosenshine, 1977; Rosenshine & Berliner, 1978).* Direct instruction includes the opportunity variable and also refers to a syndrome of classroom variables that have empirical underpinnings. For example, classrooms with an academic orientation, rather than an affective orientation, achieve higher (Fisher et al., 1978b); also, classrooms in which the time spent was academically focused are found to be consistently higher in achievement. When large amounts of time were spent in story telling, art, music, or play, as opposed to reading and mathematics activities, negative correlations with achievement were found (Stallings & Kaskowitz, 1974). Related to these findings is the well-established consistent finding that those classes and schools generally called "open" or "humanistic" do not do as well on academic outcomes as those classes and schools generally regarded as "traditional." Moreover, even attitudes about self, school, and subject matter are not higher in the more open educational programs (Gage, 1978).

Academic feedback has been found to be positively associated with student learning. Academic feedback is defined as information given to the student about whether his answers were right or wrong. Many different specific behaviors were conceptualized as fulfilling this function, including answering questions in class, checking papers, programmed text, and oral reading. The percentage of instructional time during which the student received feedback was positively related to student engagement rate and to achievement. Hence, more academic feedback leads to higher engagement and achievement. (Fisher et al., 1978b).

---

*Not to be confused with the Direct Instructional Program of Becker and Engelmann (1978).

Structuring of the lesson and giving directions on task procedures have been found to be positively associated with student success rate. Teachers who gave directions more often and spent time discussing the structure of the lesson had students who showed a higher success rate. Students sometimes do not know what they are supposed to be doing or how they are supposed to mark a particular worksheet. Clarifying activities by the teacher helps raise student achievement, probably by affecting the "success rate" component of Academic Learning Time (Fisher et al., 1978b; Tikunoff, Berliner, & Rist, 1975).

The teacher behavior of monitoring was found to be important. When academic monitoring acts were high, the teacher was able to keep children engaged in their assigned tasks. This correlated positively with achievement (Fisher et al., 1978b).

The classroom environment was also found to be important. Classroom environments, characterized as cooperative on academic tasks, warm, democratic, convivial, and with high levels of student responsibility for academic work, showed up as positive predictors of achievement (Fisher et al., 1978b; Tikunoff et al., 1975).

The conclusions reached from these studies and the attempt to define direct instruction leads to a simple one-sentence statement that best summarizes what we now know: *If the tests they use are matched to the curriculum they teach, then elementary school teachers who find ways to put students into contact with the academic curriculum, and keep them in contact with that curriculum, while maintaining a convivial classroom atmosphere, are successful in promoting reading (and mathematics) achievement.*

The "learning student" in such classrooms can then be described. First, she or he works on an academic task that is designed to result in increased knowledge or skills. We have noted that the amount of time that the student spends in a given knowledge or skill area is directly and positively related to learning in the same area. Furthermore, this appears to be as true for the more conceptual knowledge areas as it is for the more basic skill areas. For example, comprehension skills are as highly related to time spent in working on comprehension as are decoding skills to time in decoding. Therefore, the learning student spends relatively great amounts of time working on tasks that are directly related to the subject matter to be learned and the test used to measure that learning.

The learning student is also very attentive. He or she is actively involved in the task at hand, probably with some enthusiasm. The learning student is busy performing the academic part of the task, rather than sharpening pencils, looking for a book, or waiting in line to ask the teacher a question. She or he is not "socializing" or daydreaming. Nevertheless, the student is enjoying the activity. Paying attention for relatively long periods of time does not upset the student. Furthermore, his or her success on the task makes this active involvement more enjoyable.

The learning student spends a lot of time practicing and reviewing skills. She or he undertakes an activity related to a new skill only after thoroughly learning skills prerequisite to the new skill, so that she or he virtually never encounters an activity that is really entirely "new." At a younger age and/or an earlier stage of education, the learning student spends relatively greater periods of time practicing and reviewing. As she or he advances academically, it becomes possible to reduce the proportion of time spent practicing and reviewing. There is always some need for consolidation of acquired skills (practice), but as the student advances she or he actually "learns how to learn," so that it becomes easier to acquire newer skills without as long a period for consolidation of prerequisite skills. In addition, attempts to learn the newer skills probably serve to consolidate the previously acquired skills, insofar as the more advanced student is applying these acquired skills to the newer skills.

The need to spend relatively great amounts of time practicing and reviewing skills is probably partially a function of the fact that students are *not* constantly tutored. That is, students spend the majority of their time working independently or with only intermittent teacher contact. Students, particularly less-advanced students, have difficulty acquiring new skills on their own, so they need to practice previously acquired skills when working alone. Therefore, as students become more advanced, they should not only be able to spend less time practicing and reviewing, but should also be able to spend more time working independently.

The "learning student" is not necessarily an unhappy student. The learning student does not learn to dislike learning. Lots of hard work may sound undesirable to most people. However, we do not find any evidence that students are less satisfied when the sheer quantity of work (allocated time) is relatively great. Furthermore, we do not find that students who pay more attention (work intensively) acquire a distaste for learning. In fact, one could assume that when attention is the result of interest and enthusiasm, rather than coercion, then attention represents a more positive attitude toward learning. Student engagement does *not* appear, generally, to be produced by the coercive demands of the teacher.

It is interesting to note that the high-success component of learning is associated with more positive student attitudes. Successful students probably enjoy learning more because of their success. Failure, even when it is only occasional, appears to result in a more negative attitude among younger students. This may be less true, however, as students acquire more academic experience and become more accustomed to school.

Discussing what the learning student looks like is done to help teachers think about what they are trying to accomplish in the classroom. To some extent, the characteristics of the learning student are under the direct control of the teacher. Teachers make decisions about what to teach and how much time to spend on a particular goal. Teachers should be aware of how much time is really being spent on different skill areas. Classroom time is limited, so teachers should be careful to spend time on those activities that they consider the most important. If some skills are particularly important for students, it would be reasonable to spend large amounts of time on those skills. Thus, the depth of coverage is important. But teachers need also to examine the breadth of their curriculum. A wide range of content usually needs to be addressed if success on achievement tests is to be noted.

Student success rate is also largely under the direct control of the teacher. As teachers assign tasks to students, they should try to match the task to the student's skill level so as to provide frequent instructional sequences leading to high success. This strategy is particularly promising at earlier grades and for less advanced students. Note that there have been previous advocates of this approach (programmed learning and mastery learning). However, many teachers probably do not recognize the extent to which less advanced students need practice and review. The other side of the success role is also important. Teachers should always be careful to avoid giving a student a task which is extremely difficult.

Recommendations for Improving Schooling

1. Ways are needed to monitor and then to increase allocated time if it is low. The monitoring can be done by teachers, with or without colleagues. It is very revealing and sometimes very startling to teachers. The monitoring of allocated time leads to a confrontation with the single most important issue facing teachers, school districts, and state agencies: What is to be taught? More of one thing means, in a finite system, less of another. Those objectives of education receiving high priority by society ought to be the objectives receiving emphasis in the curriculum within the classroom. These objectives should be allocated more time, while other less important objectives should receive less allocated time. This very obvious recommendation is not new. *But it is not now being implemented.* The lack of implementation is not always conscious. Many teachers are never given feedback about these aspects of their teaching. They can, therefore, end up building some perfectly constructed teaching-learning units that would have a very low time priority in the community they serve. In what ways can the classroom walls be scaled and feedback given to teachers about the use of time? What district resources can be mustered to provide this

kind of monitoring? What professional rights of teachers must be taken into account in order to do such monitoring?

2. An intensive examination of the school curriculum and the content of examinations is needed. The content of instruction and the content of the tests used to assess instruction should reflect those things valued by a community (see above). Moreover, the tests and the instructional activities should be congruent. It is not likely that a second grade class with no school based training in syllabication will succeed on the half dozen items measuring knowledge of syllabication that are given in the end-of-year achievement test. Who addresses these issues in each district? Who monitors whether congruence exists between what is taught and what is tested?

3. Success rate in the chosen curriculum should be high, particularly for younger and less academically oriented students. With many classes averaging 30 students, and aides often not permitted to work in instructional roles, the ability of a teacher to accurately assess the success rate for a particular child with a particular set of materials is limited. Classroom arrangements such as smaller class size, more instructional aides, more time allocated per day to assess class work, quicker ways to provide help when students are working independently, etc., all could help keep success rates high.

4. Academic Learning Time—engaged time with academic materials or activities, that are matched to some test and provide a high success rate—is a proxy for end-of-unit or end-of-year achievement. Until now, few classroom observers knew what to look for when visiting classrooms to monitor instruction. Thus, such silly notes about the grooming of the instructor and the neatness of the bulletin board would become part of an observational record. But ALT, with known relations to achievement, can be monitored at any time; ALT may be considered learning, *as it occurs*. This means that the end-of-year achievement test need no longer be the sole criteria of successful teaching. Successful teaching can now be defined, in part, although ALT is a process variable. This is a very important and new conception of teaching and learning.

5. The academic press and other environment variables in a class are important. No one denies that schooling is to socialize students, to explore feelings, to learn the arts, etc. But when 76 minutes per day are spent in transitions from activity to activity (real data), or 85 percent of the elementary school day is spent in other than reading and mathematics activities (real data), young students cannot take very seriously the importance of academic achievement. These variables can be monitored and simple suggestions can be provided for improvement.

6. Teaching functions, rather than teaching methods or skills, are important. "No one technique of instruction is clearly associated with disastrous outcomes or successful ones" (Cooley & Leinhardt, 1978, p. 40). McDonald and Elias (1976), in related research, found that there were patterns of instruction that were effective, but these differed from teacher to teacher, from grade to grade, and from context to context. Certain teaching functions, however, always need to be met for successful classroom experiences to occur. Academic monitoring must be accomplished in some form, though many different teaching behaviors can fulfill the function. Diagnostic and prescriptive functions must be carried out, though these can be accomplished in many different ways and can be done well or poorly, in individualized or non-individualized settings. Nonetheless, some diagnosis and prescription must take place. Feedback must be given to learners. Dozens of ways to provide such feedback are known. What is important is that feedback in some form takes place. We have become too concerned about the relative effectiveness of open classrooms, personalized instruction, deductive and inductive methods, high and low cognitive levels of questions, etc., and have lost sight of the fact that certain teaching functions must be met regardless of the method used. There probably is no one best way to teach anything to all students, but there probably are similarities in the teaching functions met in all successful ways of teaching.

7. Because of the complexities of today's elementary school classroom a conception of the teacher as executive is important in training teachers. Today's teacher is not trained to

manage four resource centers, traveling students, the scheduling of speech pathologists, special educators, aides, paraprofessionals, computer terminal time, and also engage in record keeping for mainstreamed students and the preparation of individualized programs for all students. To find time for planning and carrying out direct instructional activity requires executive skill. But training for executive behavior is noticeably lacking in both preservice and inservice teacher education programs.

8. The use of most games, movies, slide-tape materials, television shows, etc., regardless of their purported educational value, must be treated with suspicion. Heavy investment in technological aids of any kind is suspect because the match of what those materials teach with both the tests used to assess instruction and the accepted curriculum is usually quite poor. Most students' involvement in games or with media can be defended, but *not* by relying on arguments about their *direct* instructional effects. In fact, empirical evidence on the relation between time spent in such activities and achievement is negative.

Conclusions

Simple conclusions about what is successful in classroom teaching and learning have been provided. After many millions of dollars spent on research, the research community has validated some commonsense notions about education. *But common sense is not common practice*, as a visit to the schools will reveal. The ways to change practice to maximize opportunity to learn are many and varied. Successful teaching and learning can occur within almost all philosophical positions and educational programs. Direct instruction, with high rates of ALT, can be obtained without turning schools into factories and without using authoritarian or coercive measures. A student's ALT falls off when classrooms are too casual and non-directive, and ALT also falls off when the schools are too coercive and authoritarian. A middle ground is needed. A convivial, democratic, warm classroom with a teacher concerned about direct instruction seems to succeed in promoting achievement of the kind measured by most standardized achievement tests.

References

BECKER, W.C., & ENGELMANN, S. The direct instruction model. In R. Rhine (Ed.), *Encouraging change in America's schools: A decade of experimentation.* New York: Academic Press, 1978.

BERLINER, D.C., & ROSENSHINE, B.V. The acquisition of knowledge in the classroom. In R.C. Anderson, R.J. Spiro, & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge.* Hillsdale, N.J.: Erlbaum, 1977.

BORG, W.R. Teacher coverage of academic content and pupil achievement. *Journal of Educational Psychology.* Manuscript submitted for publication, 1978.

COOLEY, W.W., & LEINHARDT, G. *The instructional dimensions study* (Final Report). Pittsburgh: Learning Research and Development Center, University of Pittsburgh, August 1978.

DISHAW, M. *Descriptions of allocated time to content areas for the A-B period* (Beginning Teacher Evaluation Study Tech. Note IV-11a). San Francisco: Far West Laboratory for Educational Research and Development, 1977. (a)

DISHAW, M. *Descriptions of allocated time to content areas for the B-C period* (Beginning Teacher Evaluation Study Tech. Note IV-11b). San Francisco: Far West Laboratory for Educational Research and Development, 1977. (b)

FILBY, N.N., & MARLIAVE, R.N. *Descriptions of distributions of ALT within and across classes during the A-B period* (Beginning Teacher Evaluation Study Tech. Note IV-1a). San Francisco: Far West Laboratory for Educational Research and Development, 1977.

FISHER, C.W., BERLINER, D.C., FILBY, N.N., MARLIAVE, R., CAHEN, L.S., DISHAW, M.M, & MOORE, J.E. *Teaching and learning in the elementary school; A summary of the Beginning Teacher Evaluation Study* (BTES Rep. VII-1). San Francisco: Far West Laboratory for Educational Research and Development, September 1978. (a)

FISHER, C.W., FILBY, N.N., & MARLIAVE, R.S. *Descriptions of distributions of ALT within and across classes during the B-C period* (BTES Tech. Note IV-1b). San Francisco: Far West Laboratory for Educational Research and Development, October 1977.

FISHER, C.W., FILBY, N.N., MARLIAVE, R., CAHEN, L.S., DISHAW, M.M, MOORE, J.E., & BERLINER, D.C. *Teaching behaviors, academic learning time, and student achievement* (Final Report of Phase III-B Beginning Teacher Evaluation Study). San Francisco: Far West Laboratory for Educational Research and Development, June 1978. (b)

GAGE, N.L. *The scientific basis of the art of teaching.* New York: Teachers College Press, 1978.

MARLIAVE, R.N., FISHER, C.W., & FILBY, N.N. *Alternative procedures for collecting instructional time data: When can you ask the teacher and when must you observe for yourself?* Paper presented at the meetings of the American Educational Research Association, New York City, April 4-8, 1977.

MCDONALD, F.J., & ELIAS, P. *Beginning Teacher Evaluation Study, Phase II technical summary, final report.* Princeton, N.J.: Educational Testing Service, 1976.

ROSENSHINE, B.V., & BERLINER, D.C. Academic engaged time. *British Journal of Teacher Education*, 1978, *4*, 3-16.

STALLINGS, J.A., & KASKOWITZ, D. *Follow-through classroom observation evaluation, 1972-73.* Menlo Park, Calif.: Stanford Research Institute, 1974.

TIKUNOFF, W.J., BERLINER, D.C., & RIST, R.C. *Special Study A: An ethnographic study of the forty classrooms of the Beginning Teacher Evaluation Study known sample.* San Francisco: Far West Laboratory for Educational Research and Development, 1975.

The Role of Reading in Bilingual Contexts

Warwick B. Elley
University of the South Pacific

In spite of the rapid upsurge of interest in bilingual education in the past decade, we are forced to agree with Laosa (1978) that "comparatively little research has been conducted on bilingualism and bilingual education," and "the research literature that does exist is fraught with contradictory results." Under such circumstances, instructional methods are adopted and abandoned arbitrarily. Extreme and contrary views are widely expressed on how and when we should teach children second languages, and there is uncertainty amongst researchers about which variables we should be studying.

It is recorded that 16 percent of American children now have English as their second language. In many parts of the world the figure is considerably higher. In the South Pacific, where the writer is currently working, more than 95 percent of the school children learn English as a second or foreign language. For these children, the uncertainty which reigns about acceptable theories and practices in learning English in schools is critical. For English is the language of the high school and the university and for most it is the path to secure employment. Without fluency in English, the South Pacific child stands little chance of succeeding either in his school examination or in a white-collar job. But without fluency in his own language, he loses contact with his family and friends, and the traditions that enhance his self-respect. The position is similar in many parts of the American Territories of the Pacific, in the Philippines, in Africa, in South

East Asia, and in many more countries where the language of the home diverges from that of the school. How can children best acquire the language of two cultures? If we are to help the cause of millions of children in bilingual settings, there is an urgent need for clarification of the issues involved in learning a second language.

In this article, it is planned to examine some of these issues concerning the best way to teach English as a second or foreign language, to describe some data which attempts to throw light on these problems, and to propose a new and increased role for reading in improving instruction in English as a second language. For it is argued here that one major reason for the slow progress of most children in bilingual education is the deliberate neglect of reading in this process. The current ferment in psycholinguistics and in research into the process of reading comprehension has produced new and important insights into the nature of language acquisition. Unfortunately, little of this knowledge has rubbed off in bilingual education. With so many futures at stake, we cannot afford to leave these gaps unclosed.

The Bilingual Context

In this article, bilingual education refers to education which is provided through the medium of two languages. This includes all cases where children come to school speaking a vernacular and switch at some stage of their schooling to a second (or third) language as the medium of instruction, while retaining and/or developing their competence in the first. In some circumstances the transition takes place immediately upon entrance to school, as in the French immersion programs pioneered by Lambert (1972). However, the typical pattern in the South Pacific is for children to arrive at school with a good oral command of their mother tongue—be it Fijian, Hindi, Samoan, Tongan, or one of the many vernaculars in current use—and to have most of their instruction in the first three or four years in that vernacular. The switch is gradual, with no more than 30 minutes a day of English as a second language in the first year, and a marked increase in English in grades three or four. At this stage most instruction is in English in all subjects, apart from a lesson a day in the vernacular, up to the middle of high school. In countries which have only one vernacular, the switch takes place later; in countries with two or three widely used vernaculars the transition is largely completed by class four; in countries which have a multitude of home languages, the switch often occurs as soon as the child begins school. In most cases the teachers are bilingual, and moving back and forth from one language to the other in one lesson is very common. The basis for these various policies is found in a mixture of tradition, convenience, and parental pressures. Rarely can the administrator point to experimental evidence, or even an assessment of needs.

Approaches to Language Learning in Bilingual Contexts

Any reader professing an interest in methodology of second language instruction could hardly fail to recognize dramatic changes in the methods and rationales of such instruction in recent years. As Strevens (1977) points out, "Unlike the scientific disciplines of linguistics and psychology, with which it has been linked in the past two decades, language teaching has remained an art and craft." And like all arts, it is subject to rapid changes of fashion.

The prevailing fashion in foreign language teaching early in this century was formal and classical, with a heavy emphasis on grammar and translation. During the 1920s the direct method was introduced in some countries. Stress was placed on hearing the target language spoken, much as the child learns his first language naturally, and the learner's native tongue was not to be used. The idea of strict controls of vocabulary and structures was popularized in the 1930s, so that the learner's input at the point of instruction was to be carefully matched to his knowledge and stage of readiness. Audiolingual methods were widely used in the 1940s, when crash courses were required for armed service personnel, particularly in Britain and the United States. These methods were further developed in the 1950s, drawing on theoretical rationales from the work of such linguists as Bloomfield and Fries, and of behaviorist psychologists. Language was seen as a set of habits to be learned and drilled like any motor skills, under conditions of repetition and reinforcement. The rules and structures of the target language were to be taught one at a time, orally at first, later through practice in reading and writing. In the developing countries of Africa, Asia, and the South Pacific where English was taught as a second or foreign language to young children, alongside or after the vernacular, it has been the audiolingual method which commanded the most adherents.

More recently the work of Chomsky and the "nativists" of linguistics has given rise to a new approach to the learning of languages, both first and second. Chomsky's attack on the behaviorist view that languages are learned by imitation, reinforcement, generalization, and other such concepts has been widely publicized. Recent empirical study has provided much evidence of the young child's remarkable ability to produce unheard language structures and to generate novel ones—abilities which seemed to suggest a child who is predisposed to acquire certain patterns in language learning, to hypothesize and to correct his own hypotheses, and thus to gradually approximate the language of the adult. The key concept here is the child's intuitive creation of his own language code, not the conscious, systematic block-building analogy of the structuralist-behaviorist approach. It is as if the child is born with a predisposition to acquire certain language forms. All the teacher needs to do is to expose the child to a variety of structures and vocabulary in meaningful communication, and the child will develop his own language rules, easily and intuitively.

Many more methods and theories have been promulgated in the search for the best or optimum method of teaching a second language. Kennedy (1973) lists 13 different methods which have been seriously used and advocated, and suggests that all have been partially successful. However, his point should be reiterated that nearly half the world's children may be fluent in two or more languages without formal instruction; yet few taught formally, in school, have ever been very fluent with any of the methods he enumerates. Clearly we have much to learn about how best to teach another language.

Why Is Second Language Learning Ineffective?

Why do second language learners find their task so difficult? The reasons advanced for Kennedy's arresting conclusion have been many and various. Some point to the difference in motivation. The child who has mastered one language, has already learned how to satisfy his natural urge to communicate with those around him. Learning the second language for a distant examination, or vocational purpose, or to satisfy a bureaucratic or academic requirement, is a less compelling reason than the ever-present desire to communicate with friends and family. As MacNamara (1971) has pointed out, the teacher rarely has anything important to say to the child, and the child has nothing important to say to the teacher.

Others point to the interference effects of the first language on the second. Contrastive analysis has indeed identified a number of deviations in the expressions of second language learners—in phonology and syntax—which are readily traceable to the effects of the first.

However, more recent studies (Dulay & Burt 1973) suggest that these interference effects, at least in the structures of the second language, are limited to less than 10 percent of observed deviations, while others argue that these effects are limited to the first stages of learning, and appear subsequently only when the speaker is tired or under stress. These hypotheses need further investigation in a variety of language situations.

Another plausible reason for the difficulties of the second language learner is the limited opportunity to hear and practice the language. Even in the "immersion" programs where pupils are required to use the target language at all times, they spend only 25 hours per week in the second language environment. And in many parts of the South Pacific the figure is considerably less than this. By contrast, the first language learner is exposed for most of his waking hours—over 100 hours per week.

Many theorists regard imitation or modelling as a critical feature in learning a second or foreign language. When the teaching model is sound, where the structures are substantially correct, the vocabulary judiciously chosen, the pronunciation accurate, the message clear and important—under these circumstances the child has a head start over one exposed to poor teaching models. Regrettably, bilingual teachers are not often equally fluent in their exposition of both languages. Examples abound of clumsy expressions, poor spelling and faulty diction. Townsend and Homer (1978) quote an example from an English lesson in a Pacific elementary school, where the teacher was struggling to teach her children that "If we doesn't be careful with these matches, it will burn ourselves." Where children and parents recognize their teachers' inadequacies, as they frequently do, progress is inevitably slow.

Another hypothesis which may explain the difficulties experienced by the young second language learner is the artificiality of the language he is exposed to. Rarely do the drills and patterns practiced in early oral and written language sessions approximate the natural language that the first language learner is exposed to. As Kennedy (1973) puts it: "From the incredible structural richness of a language, we, the teachers, select phonological, syntactic, lexical and thematic items; we decide and arrange the sequence of their presentation to the student; we force him to practice the rules we think are being learned" (p. 75).

Certainly, common sense suggests that we make some allowance for simplifying language for young learners. If however, the child can learn a first language without systematic screening out of new words and structures, it behooves us to keep an open mind on the question of how far this screening should be taken with second language learners. If children do learn a language by deducing their own rules about its structure and occasions for use, the systematic screening process may actually slow down the learning. When the child never has the chance to hypothesize, to check, and to confirm or correct, his pace of learning will inevitably be impeded. This question of the optimum ration of known to unknown words is still an open one.

In sum, the second language learner, for a variety of reasons, has an uphill battle ahead of him, and instructional approaches have, to date, rarely succeeded in producing fluent confident users of the second or foreign language. Clearly we need some fresh approaches to the problem. If present day methods are uninspiring, it is time to examine their assumptions and see whether they are at fault.

The Audiolingual Method of Instruction

Writing in 1975 Diller pointed out that 15 years earlier, in a "state of the art" address to the Ninth International Congress on Linguistics, William Moulton had been able to summarize two decades of increasing consensus amongst his peers, that language was best taught through an audiolingual approach, which consisted of "mimicry, memorization and pattern drill." Language was not thought of as rule-governed, but as a set of speech habits learned by conditioning and drill.

It is true that the audiolingual approach was the widely accepted method of TESL in the early 60s. Structuralism in linguistics was at its height; and few linguists had any substantial grounds for challenging the psychological tradition that languages are skills, learned chiefly by the mimic-model, with abundant practice and drills.

Diller went on to assert, however, that the linguists of the middle 1970s saw no such consensus in the methodology of language teaching. Adherents of the audiolingual method are still alive and well, but their case is now widely challenged. Let us examine a typical audiolingual program and its assumptions.

In the South Pacific, the audiolingual tradition has long dominated the English language learning scheme. During the early 1960s, Gloria Tate developed a program for use with Cook Islands Maori children learning English in elementary schools, using structuralist principles derived mainly from those of Charles Fries, and a behaviorist orientation in learning. This program has subsequently been extended for use in all countries of the South Pacific, as well as many of the American territories. It is now produced and distributed by the South Pacific Commission (SPC) and is used in virtually all primary schools in the region. Its principles are taught in all teacher training colleges and promoted by the advisers associated with all education departments.

The rationale of the SPC/Tate Program has been set out in a number of handbooks (Tate 1967, 1971). Briefly, it assumes that language is a set of habits learned by imitating good models through regular drills. Each new structure and word is to be consciously identified and its use practiced in oral form, following the teacher's model. Thus the scheme is predominantly an oral English program. Subsequently, the same structures and words are practiced in reading through the specially written SPC Junior Reader Series and several other sets of graded readers. Errors are to be avoided at all costs, as they are difficult to eradicate. Therefore particular care is taken to ensure that all English language the children hear or see is strictly controlled, so that errors do not occur. One way to ensure this avoidance of error is to so control the reading books associated with the oral program that children are never exposed in print to structures and vocabulary that they have not previously learned in their oral lessons. Indeed, the guiding principle is that there should be a gap of twelve months between the presentation of oral and written forms in the early stages, narrowing to a three month gap or less at the upper primary school level.

These principles are not unusual in an audiolingual program. More specifically, the idea that mastery of the oral language should take place before exposure to written forms, in case of interference, is still widely recommended. In his authoritative summary of bilingual teaching in the United Kingdom, Derrick Sharp (1973) maintains that second language learning should begin in the infant school (age five) and should remain an entirely oral activity until the late junior stages (ages ten to eleven) or the early secondary years for the less able pupils. He refers also to the progress made in recent years in the stricter control over vocabulary and the carefully graded readers pupils are exposed to. Similarly, an orthodox view in North America is summed up by Ching (1976) when she says that "before bilingual children can learn to read English, they must be able to understand and speak it effectively," (p. 4) and "items should be presented in spoken form before they are presented in written form" (p. 33). It is important to examine this assumption.

Why Delay Reading?

The reasons for the audio-lingual claim that reading should be so long delayed, and thus relegated to a minor role in language development, require further investigation. They clearly have inhibiting effects on teachers and learners. In the face of this belief, pupils are not encouraged to read widely. Teachers are not active in building up libraries. Parents can have little role to play in leading children to books. Is this restriction on reading really warranted?

Wheeler (1974) suggests that the delay is based on a false analogy with first language learners. Monoglot children normally learn to read only when they have built up a large oral language repertoire in their preschool years. The fund of knowledge about words and structures certainly does provide a useful source of semantic cues for speeding up the reading process when other cues are weak. But the learner can also resort to cues from pictures, from text redundancy, from analogy with the vernacular, or from the teacher's translation. To wait until the children have acquired a large repertoire means a long wait, during which they are excessively dependent on oral methods. Many children are "visualizers," who prefer to see what they are learning in print. Certainly there are empirical data which show high correlation between oral language fluency and early reading achievement, but we cannot argue from this fact that the oral strength is the cause of the facility in reading. In fact, after a longitudinal study of the progress of Samoan and Maori children in learning to read, Clay (1970) suggested that early reading progress depended more on "progress in visual perception of print" than on sophisticated command of oral English.

Another allegation made against the early introduction of the printed word to the young bilingual learner is that there is a danger of interference. The child who sees a printed form which does not fit phonetically with his pronunciation of it is likely to revert to a speech pattern which fits his vernacular preconception of how it is pronounced. Evidence for this view is not presented by its adherents, however, and it is not likely to occur in cases where they are considerably different. Even if it were a danger, a case could still be made for the simultaneous introduction of both oral and written forms, rather than a lengthy delay. Of course, the argument could also be turned around. Children are just as likely to invent incorrect spellings for the oral forms they have learned if they are not given the correct visual form of the word. And this will interfere with later written composition work. But it is time to look at the empirical findings.

Research on the Use of Print in Learning Language

It has long been recommended by educational psychologists that a multisensory approach to learning new material will provide better results than a single-sensory method. When new spelling words are taught systematically, it is common procedure to have children see the new word, say it, write it on paper, write it in the air, close their eyes and visualize it, and thus produce a greater impact on the pupil's brain than a single oral or visual presentation might. Does this multisensory principle apply to learning a new language? According to research on bilingual education in Wales, it certainly does.

Dodson (1967) compared four methods of learning sentences in a second language on the part of pupils of both elementary (ages eight to nine) and secondary school (ages thirteen to fourteen). The four variables he studied were the presentation of the *spoken* form of the target sentences, the *written* form, an accompanying *picture*, and the *mother tongue* equivalent. Each child was interviewed individually, and the criteria tested were fluency in speech and comprehension of meaning. For the younger children the most effective combination, by far, was that in which the picture, the spoken word, and the printed word were presented to the child. Least efficient was the picture plus spoken word. The presentation of the mother tongue equivalent assisted learning for most children, but not as much as the presentation of the printed sentences. The trends were similar for the secondary school pupils, but the value of vernacular equivalents appeared to be nearly as great as that of the printed word. On a time-to-learn criterion, the best approach was that which employed all four variables simultaneously. Clearly the multi-sensory approach finds support here.

In a series of follow-up studies in classroom contexts investigating the role of the printed word in learning German sentences, Dodson found similar results. When the printed words were exposed while the children listened to and repeated the teacher's oral presentation, the teaching time required was greatly reduced. Further confirmation was found in the learning of number concepts and in the recall of sentences learned earlier. In each case the printed word was a clearly effective aid to learning.

As for the problem of interference in pronunciation due to the presentation of the words in print, Dodson states that it hardly ever arose. "Only a handful of letter combinations caused temporary interference lasting no longer than a few minutes, before a correct pronunciation was achieved" (1967, p. 20). On the other hand, the absence of the printed word was a problem for many who were unable "to sort out the tangle where one word ended and the next began." In Dodson's view, the presence of the printed word enables them to match particular sounds with initial letters of the words—or syllables in polysyllabic words—and so provide signposts to help them through the language. It also led to fewer problems of spelling later, and provided more time to practice good pronunciation.

As Dodson pointed out, the use of print in initial language learning was unpopular with theorists at the time of his research. Whatever arguments prompted this viewpoint seem to be consistently demolished by Dodson's experiments. Yet, over a decade later the assumption is still widely held, and its implications extensively preached to teachers and would-be teachers.

Can Second Language Learners Teach Themselves?

If children do not suffer when exposed to print while learning oral language, the question must now be put as to whether they might actually benefit by learning new language from the printed word, independently without the teacher's guidance. Is there learning potential in unfamiliar words? To what extent can pupils cope with new structures in print? Can they make some sense out of unknown language and go on to teach themselves new forms?

The audio-lingual approach is again adamant on this point. Reading is to follow, not lead. Tate (1971) points out that "children should understand the meaning of whole printed sentences and of whole printed paragraphs immediately when they read them." Pitman (1974) in his rationale for the SPC/Tate Program asserts that: "language cannot be taught through a reading program."

These views stand in marked contrast to those expressed by recent psycholinguistic theories of reading in a first language. The audio-lingual proponent says we should avoid error at all costs. Yet first language learners who are prepared to guess, to hypothesize and confirm are now considered to be the best language learners. Goodman sees reading as a psycholinguistic guessing game. "Reading becomes a sample, predict, test, confirm and correct-when-necessary approach" (Goodman 1976, p. 238). Frank Smith confirms this viewpoint. "The most preferred and efficient strategies for proficient readers when they come across a word that is unfamiliar are to skip or to predict from context or by analogy with other words. By conducting experiments as we read, not only do we learn to recognize new words, we learn everything else to do with reading" (Smith 1978, p. 97). "Children do more than learn to read through reading; they learn language" (p. 70).

To what extent are these principles applicable to second language learners? Can a child without an extensive fund of words and structures be expected to detect redundancy in English, to learn from context? If we can venture to expose him to the unfamiliar, without harm, then many practical implications will follow.

Firstly, more interesting reading material can be put into children's hands, materials that are not so strictly graded that high interest (but unfamiliar) words are screened out. Reading could become a more attractive activity. In consequence, pupils' motivation to attribute meaning to new words will increase. They will want to know in order to satisfy natural curiosity, not to satisfy the teacher. On the other hand, those who are perpetually protected from the unfamiliar printed word will not develop a set to predict or hypothesize. Faced with a new word or structure they will stop and ask for help. Those who never practice independence will not achieve it. Furthermore, if children who read widely learn more about their language, then the sooner they are encouraged to forge ahead on their own the better. To follow the audio-lingual restrictions is to discourage extensive reading for interest, to slow down growth towards independence in learning and to retard language development.

What evidence is there on the ability of second language learners to cope with new language structures in print? On the basis of observation of children in Hawaii and California and with the help of videotape playbacks, Joan Rubin (1975) has isolated some of the strategies characteristic of good second language learners. At the top of her list is the guessing strategy. "The good guesser uses his feel for grammatical structures, clues from the lexical items he recognizes, clues from the redundancy in the mesage. He uses nonverbal cues, word association clues, outside knowledge (his general knowledge of society, of similarities to his native language). He makes inferences as to the purpose, intent, point of view of a message." Rubin sees no difference between first and second language learners in these respects. Nor does Twaddell (1973) or Vivian Cook (1969). And guessing is possible for children as well as adults. Yet guessing is strongly discouraged by the audio-lingual approach. Can we find some evidence on this point? Perhaps it does not apply with young children. Perhaps there are qualifications in situations where English is not used outside the classroom. What kind of structures can children cope with? Can they use context to guess the meaning of unfamiliar forms? These questions are important in portraying a role for reading in bilingual contexts.

Experiment on Learning from Context in Print

In an attempt to probe some of these issues, the writer conducted an investigation with bilingual children in grade five in two South Pacific countries. In one country, Fiji, English is learned predominantly as a second language, after initial reading instruction in the vernacular. In the other country, Niue, English has the role of a foreign language, as it is spoken only rarely outside the school. In both countries oral English commences in grade one; English reading, in grade three.

Twenty-four structures taught orally in the Tate Oral Program in grade six were identified and placed in the context of meaningful sentences or short paragraphs. These structures had not been taught to the pupils in the investigation. There is a rigidity about the sequence of the Tate Syllabus which ensures uniformity in this respect. The question to be examined was how well the pupils would comprehend each of the structures in a meaningful reading context, *before* they had learned them orally.

The sample of pupils studied consisted of 100 typical grade five children (eleven years old) from three schools. One was a below-average urban school of Fijian and Indian children from a predominantly working-class suburb of Suva; another was a middle-class school in an innercity suburb, with a predominance of Indian pupils, again in Suva. The third class was drawn from Niue, in the small town of Alofi. All pupils had learned to read first in their home language, and were in their first year of learning to read in English. All were using the Tate Oral Syllabus and its associated structured readers.

The unfamiliar structures to be read were presented in short simple sentences or paragraphs, and a parallel set of sentences containing familiar structures was prepared as a control device. Each child read twelve of the familiar and twelve of the unfamiliar structures. The study was conducted in intact classroom groups. After reading each sentence the children responded to open-ended questions, assessing their comprehension of the meaning of the structure in question.

Examples

(1) (a) *Unfamiliar structure*: The children next door *were frightened by* the little dogs.
(b) *Familiar structure:* The little dogs *frightened* the children.

QUESTIONS (i) Who was afraid? ____________________
(ii) Who frightened them? ____________________

(2) (a) *Unfamiliar structure:* Mother said, "There are *hardly any* bananas left, Rima. Go and buy some more. There's a shop *not far from here.*"

(b) *Familiar structure:* Mother said, "We have *only a few* bananas left, Rima. Go and buy some more. I saw a shop *near here.*"

QUESTIONS (i) Did Mother have any bananas at all? ____________
(ii) Was there a shop near them? ____________________

In each class, pupils were assigned at random to one of two groups, X and Y. Pupils in Group X would read sentence 1(a); pupils in Group Y would read sentence 1(b). Both groups responded to the same questions. Then pupils in Group X would read Sentence 2(b); those in Group Y would read Sentence 2(a); and so on.

Table 1 sets out the results for 18 of the comprehension questions asked, expressed in percentage form. For these 18 familiar structures, the pupils had little difficulty. All percentages were well above 60 percent and very similar to the figures for the comparable familiar structures. The mean score on the unfamiliar structures was 78.33 percent, just 4.61 percent less than the mean score for the familiar structures. Apparently these pupils can comprehend structures which they have not been specifically taught. Apparently they can and do learn from context. Yet the Tate Program prevents these children from either seeing or using such structrures.

Table 2 presents the figures for the remaining 6 structures which were not well understood. All showed percentage correct scores less than 60 percent, and the mean score was 25 percent or 53 percent less than that obtained from the familiar structures. Closer inspection suggested that these were inherently difficult language forms even for more mature children. Perhaps they had no place in the grade six program at all. To investigate this issue, the same sentences and test questions were presented to a class of 35 grade seven pupils in Suva from the same suburb as the first group of class five children. These children had been taught the

Table 1
Grade Six Language Structures Well Understood

	Unfamiliar Structures	Comparable Familiar Structures
1. In order to . . .	68%	76%
2. When Joe had done . . .	86%	96%
3. . . . were frightened by . . .	89%	84%
4. . . . can't be true.	68%	82%
5. . . . went to his mother carrying . . .	76%	82%
6. . . . on which the paper was put . . .	65%	90%
7. . . . without making a sound	79%	92%
8. It's such a big cat that . . .	92%	84%
9. . . . too strong for me to . . .	96%	95%
10. Can it be a . . .	87%	66%
11. It must be a . . .	76%	71%
12. It doesn't take him long to . . .	63%	76%
13. He is also big.	76%	94%
14. It must take him a long . . .	71%	84%
15. There's a shop not far from . . .	79%	82%
16. That's not enough for . . .	76%	76%
17. . . . cents for you to take.	87%	87%
18. . . . so that he wouldn't be	76%	76%
Mean:	78%	83%

particular structures tested in the grade six year. Therefore the hypothesis was that their ability to comprehend these structures in writing would be significantly better than that of class five children. Table 2 shows that this hypothesis was not, in fact, supported. Grade seven children scored 28 percent correct on these 6 structures, only 5 percent more than children in grade five and still well below the figures for the comparable familiar structures. Apparently, the effects of specific instruction in the structures in question have little or no effect on the ability of the pupils to comprehend them in print.

These findings seem to be consistent with the results of recent studies by Dulay and Burt (1973) with bilingual children in the United States. These investigators used comparative error analysis to show that Spanish-speaking elementary school pupils make the same kinds of errors in learning English as a second language, regardless of whether they were systematically taught the relevant structures or not. Their results were interpreted to

Table 2
Grade Six Language Structures Not Well Understood

	Gr. 5	(Control)	Gr. 7	(Control)
1. *Neither* Peni *nor* Subhas saw the cat...	53%	(96%)	65%	(100%)
2. Anjula *hasn't been* at school since...	34%	(78%)	48%	(84%)
3. We have *hardly any* bananas left...	13%	(39%)	13%	(68%)
4. *All* the children *but* Sashi have...	13%	(97%)	5%	(91%)
5. It's *too far for him* to...	8%	(66%)	18%	(83%)
6. *A mile further than*...	16%	(82%)	18%	(87%)
Mean:	23%	(76%)	28%	(86%)

mean that it is a waste of time to teach children syntax deliberately, a conclusion which is clearly at variance with the spirit of audio-lingual methods.

Such findings raise serious difficulties for any program which is based on the assumption that children should be taught language structures in a particular sequence, and that reading materials should be graded in such a way that children are not exposed to the unfamiliar. On the contrary, bilingual learners in typical schools can and do learn from written context. They can make sensible and correct predictions about unfamiliar structures in meaningful settings, and instruction in such structures is apparently not helpful for them. The results in the four South Pacific schools in the study were virtually identical. The same 18 structures were well understood, regardless of whether the children were learning in a second or a foriegn language situation.

One implication of these studies is that children should be encouraged to read more extensively. If bilingual pupils can teach themselves the meaning of new language when they see it in print, then they should not be restricted to the few highly structured readers that make up the reading fare of the typical child learning in an audio-lingual program. In the South Pacific this conclusion implies a complete change of policy. For most primary schools have few or no library books; no indigenous literature exists for South Pacific children; and most pupils never read independently for pleasure. In fact, many teachers take so seriously the restriction that children should not be exposed to unfamiliar words or structures that they never read stories to children, in case they may hear something new which would cause confusion and error. Such inhibiting policies may well be the root cause of the now widely documented weaknesses in children's learning of English language in the South Pacific (see Elley, 1979; Elley & Mangubhai, 1979; Stamp, 1979; Townsend & Homer, 1978). Standards of English reading and writing are poor and apparently declining; high school students cannot understand their textbooks; university students are requesting assistance with the comprehension of lectures and texts. And where English is a problem, learning is slow in all subjects.

Evaluation of a Reading-Based Program

Not all South Pacific educators have been pleased with the audio-lingual approach to reading. The director of education in the small island of Niue, De'Ath (1978), a former New Zealand adviser on reading, was not impressed with the progress of his primary school pupils in English language, or with the lack of emphasis on reading in his pupils' English syllabus. The SPC/Tate Junior Readers, which are used in Niue and in most South Pacific schools, are deliberately graded and articulated with the oral English syllabus in such a way that pupils see and practice only familiar words and structures. The choice of language is dictated by the need for repetition of logically sequenced structures; interest, meaning, and natural language patterns are frequently sacrificed for the demands of form and structure. Typical quotations from the Junior Series Book 3 demonstrate the kinds of problems faced by the authors in making the materials appealing for children.

1. Tom has a tin and Peter has a tin, too.
   Tom is holding his tin and Peter is holding his tin, too.
2. Here are Peter and Mary.
   Are they going home?
   No, they aren't.
   They're going to Tom's house.
   Are they good children?
   Yes, they are.

Is Mary's dress clean?
Yes, it is.
Are her hands clean?
No, they aren't.

It is difficult to imagine South Pacific children reading such material for sheer enjoyment. Faced with a situation in Niue where English language was used very little outside the classroom, and the teacher's models of English inside the classroom were often at fault, De'Ath (1978) argued that the children would learn English more readily if they were provided with a richer language environment in print. While accepting the oral component of the Tate Syllabus, he questioned the limited role given the printed word for teaching new langauge, and set out to develop a new set of readers, known subsequently as the "Fiafia" Reading Books. "Fiafia" in the Niuean language means "happy."

The distinctive qualities of this new program are as follows:

1. The materials are all high interest stories, based on local people and familiar situations. (The school bus breaks down; the farmers' pigs escape and get into mischief.)
2. Early language was controlled to a certain extent with regular repetition, but unusual or "interest words" and new structures were frequently introduced to enhance the story.
3. Each book is well illustrated, usually with a touch of humor.
4. Most of the books are short enough to be read in one sitting.
5. Large supplementary readers are used in a "Shared Book Experience" method.

In this method, the teacher has the children sit around her in a group, on a mat. She presents an interesting story, "blown-up" in big print. First, she reads the story to the children; then they read it together. They stop regularly and talk about the pictures—just like Mother reading bedtime story. Then the children read the story again. The teacher calls on individual pupils to read specific words or phrases, or to answer questions about the meaning of the story. A high premium is put on group discussion, which arises naturally if there is a good story. Thus the children's own language experience can be complemented and extended. The children enjoy reading the stories together, the unison chorus approach encourages natural intonation patterns, and there is a valuable opportunity to learn basic concepts about books and book language, and to acquire new words in their context by listening to the teacher or to other children. They learn easily, without the embarrassment that comes when the teacher identifies and exposes their ignorance. This "Shared Book Experience," sometimes called the "Cooperative Book Method," is a compromise between reading *to* the pupils and reading *by* the pupils. Rather, the teacher reads *with* the pupils—a method very popular in New Zealand schools at many levels.

So, the main differences between the new Fiafia Program and the traditional SPC reading program are that the Fiafia Program depends on a good story, provoking group discussion about it, and the learning of new words and structures in the context of that story. New language is learned only in context—and a written context at that.

The SPC readers, on the other hand, are designed to help pupils practice particular words and structures. The stories are carefully graded to keep out anything not already learned in the separate oral lessons; consequently, they lack excitement. The children work gradually and systematically through the books, two pages a day, much as they might work through a mathematics book.

The situation in Niue provided an ideal context for examining the role of reading in the program of English as a foreign language. It was possible to compare a complete age group of children in grade three (eight years) who had taken the Tate Oral English Syllabus in association with the Tate Junior Reading Series in 1977 with another grade three group who had taken the same oral syllabus in association with the Fiafia Reading Program in 1978. All other instructional and home background factors were held as constant as possible. The children in both cohorts were the same age; they had been learning oral English for two years; they had learned to read in their vernacular; the hours of formal instruction in English and the vernacular were the same; the teachers involved in the two grade three programs were the same; and checks were made on reading performance of pupils at the next grade level to ensure that there was no systematic growth in English due to unknown community or school influences.

The experimental context was ideal for evaluating the effectiveness of these two approaches to English teaching—the audio-lingual Tate Program, with its minimal role for reading, and the Fiafia Program, with its major role for reading. For in the virtual absence of English outside the school, the activities of the classroom would show their effects on children's language growth in far bolder relief than would be the case where pupils learned both inside and outside the classroom. Instruction in a foreign language provides a better laboratory context to examine the influence of curriculum variables which are so often neutralized by the effects of home and community.

Design of the Evaluation

The design of the evaluation called for the testing of all Niue pupils early in their grade four year. The Tate Program pupils were assessed in March 1978; the Fiafia Program pupils in March 1979. All grade four pupils on the island were tested in both years, but the children of one of the six schools were omitted from the analysis, as that school had been the scene of all the trials of the new Program in 1977. Consequently, its pupils had not taken the Tate Program in that year.

Three tests were devised to assess the pupils' performance in English—*Reading Comprehension, Word Recognition* and *Oral Sentence Repetition.* The Reading Comprehension test was adapted from a multiple-choice sentence completion test designed for Cook Islands Maori children in an earlier project. It was trialed also in Fiji and Niue (on pupils in the trial school) and revised to form a 35-item, group-administered test with simple vocabulary and a local South Pacific flavor. The choice of vocabulary was influenced by a Niue word frequency list prepared by the author in an earlier project. It was not influenced by the contents of either the Tate or Fiafia Readers. All pupils were given this test in both years. It was administered in all schools by a Niuean education adviser, using both English and Niuean instructions and in the presence of the writer.

The Word Recognition test consisted of 50 words, graded for difficulty, which children were required to read aloud to the writer (or his Niuean counterpart). Half of the words were drawn from the Tate Readers and half from the Fiafia Readers. The first words were simple high frequency words found in both programs (hop, look, me, not, and). The last words were more difficult but still included in the Tate Syllabus at higher levels (weigh, machine, breathe, lightning). Half of the pupils, chosen at random, were administered this test in each year. The correlation between the results of the Reading Comprehension Test and the Word Recognition Test was 0.75.

The Oral Sentence Repetition Test was introduced primarily to test for incidental effects of the two reading programs. The test was adapted from one developed and validated by Clay, Gill, Glyn, McNaughton, and Solomon (1976) and consisted of 28 English sentences of increasing grammatical complexity. Each sentence was read aloud to a pupil by the writer, or his Niuean counterpart, and the pupil was required to repeat it verbatim. The rationale for such tests is that if pupils do not have a particular structure in their repertoire, they will state it incorrectly and/or omit the unfamiliar element. Recent research shows that it is a sensitive measure of language competence, rather than a test of rote memory (Hanayan, Markman, Pelletier, & Tucker, 1978). Each child, in both years, who had not taken the Word Recognition Test was given the Sentence Repetition Test by the writer or his Niuean counterpart. Three examples were given for practice. When a child scored less than 3 sentences correct in one of the 4 sections of the test, he was stopped.

Typical sentences were:
My brother's knees are dirty. (N be +)
The cat is drinking some milk. (NVN)

> Here comes a great big truck with children riding on the back. (Here VN)

The reliability figures reported by Clay et al. (1976) are over 0.90 and the correlation with the Reading Comprehension Test in this investigation was 0.54.

In addition to these three objective tests, a brief interview was conducted with each child in both years in order to gain further insight into his oral language fluency and attitudes towards reading.

Despite two teaching staff changes, for part of 1978, and a three month delay in the arrival of the supplementary readers for the Fiafia scheme, the program was taught according to plan in most cases. Some teachers made greater use of the "Shared Book Experience" than did others. The timetable was standardized, and the number of supplementary books in the school libraries was very similar from year to year.

Results of the Evaluation

Table 3 shows the mean and standard deviation for each test in both years. In each case there was a dramatic increase in mean scores under the Fiafia Program, highly significant in all tests. Thus, pupils taught with a relatively ungraded reading program, which stressed interest and meaning rather than structure, produced better reading comprehension, better word recognition, and better mastery of oral language. Furthermore, the interviews showed that both pupils and teachers enjoyed the Fiafia Program more than the Tate Readers. Gains were shown in 16 of the 18 test comparisons made, and some schools showed increases of more than 100 percent with the same teacher. Clearly, a reading-based program of this kind has much to offer. Clearly, too, the shaky assumptions which led to the audio-lingual neglect of reading are undermined even further. For after three years of English teaching under the SPC/Tate Program, the quantity of simple words recognized and understood by grade four children was little better than chance.

Table 3
Language Test Scores for Tate and Fiafia Reading Program

	Tate Program 1978			Fiafia Program 1979			
Language Test	Mean	S.D.	*N*	Mean	S.D.	*N*	Difference
Reading Comprehension	10.54	5.50	89	14.02	6.63	62	p<.001
Word Recognition	14.85	12.00	49	29.41	12.39	34	p<.001
Oral Sentence Repetition	6.51	5.00	39	10.91	6.30	31	p<.001

It is difficult to identify the particular elements which produced the marked improvements in the Fiafia Program. The 12 booklets themselves and the 36 supplementary readers are not polished examples of the publishers' craft. In fact they need thorough editing and the illustrations and photographs are often poorly produced. Nevertheless, the children were happy to read and reread them often, thus reducing the need for repetition to be built into the text.

There is a strong suggestion in the school-by-school analysis that those teachers who made greatest use of the Shared Book Experience produced the most growth; those who used it less often, produced the least. Reading specialists who promote the Shared Book Method (Nalder, 1975; Porter, 1977) would no doubt agree that it is an important ingredient, as it has convinced large numbers of New Zealand teachers who have made use of it. Ritchie (1978) found it the most effective of three programs in her evaluation of a Maori preschool project in New Zealand. It is a method which capitalizes on children's natural interest in a good story. New language is taught at the point when children are highly motivated to read. Every psychologist and practical teacher knows the benefit of intrinsic motivation.

If these explanations are plausible, the policy implications for second language learners are clear. If children can teach themselves new vocabulary and structures, in much the same way as first language learners, and they can capitalize on the Shared Book Experience in much the same way as first language learners, then it is clearly important to loosen up the restrictions we have imposed on the language environment of young children in a bilingual setting.

Conclusion

It has become obvious to many teachers in bilingual programs that there is very little transfer from patterned practice exercises to genuine oral or written communication by students. Uninspiring, repetitive drills turn students off, leaving little impact on their active language repertoire. As Ervin-Tripp (1970) points out, any learning model which concentrates on language input, without regard to selective processing by the learner, just will not work. Apparently, we cannot teach children language. We can only make it possible for them to teach themselves.

What then is the formula for better quality bilingual programs? We have seen that formal instructional programs have rarely been successful in teaching children a second language. Yet children learn first and second languages quite efficiently outside the classroom, when communication needs are genuine.

We have seen that second language learners profit best from a multisensory approach, rather than from one which neglects the printed word. We have seen, too, that typical grade five children can learn new structures in the target language, without formal tuition, if given an opportunity to do so—in the form of a meaningful context. Furthermore, we have seen that a bilingual program which stresses learning from interesting reading materials accompanied by active questioning and discussion is more effective than a formal, carefully graded reading program, tied to a tight sequence of orally taught structures.

These findings are consistent with the contentions that good language learners approach their task best when they are highly motivated to discover meanings and patterns rather than required to practice structures. If language learning is more effective when the learner is often faced with the unfamiliar and is keen to guess or hypothesize to fill the gaps, then audio-lingual programs have been guilty of retarding learning by discouraging guessing and preventing children from attempting to work out new meanings for themselves. We should encourage the teacher to reward guessing, to recognize that readers' deviations from text are frequently signs of positive thought, harmless approximations which indicate an important step towards independence in language learning.

The audio-lingual approach has been guilty, too, for neglecting the printed word. In the South Pacific, children rarely read for pleasure, and teachers rarely read to them. The few books available do not lend themselves to enjoyable reading. This state of affairs is clearly unsatisfactory.

There are probably no simple solutions to speed up learning in bilingual context. However, we could begin by attempting to exploit the potential of the printed word by building up a stock of reading materials geared to pupils' natural curiosity, their love of narrative, of excitement, of humor, their easy identification with mythical heroes and with children like themselves. We should make such factors paramount, and worry less about strictly controlled language. In this vein, extensive reading, in the shape of deliberate "book floods" in programs for bilingual Polynesian children in Auckland have shown much promise (Elley, Watson, & Cowie, 1976). Furthermore, the evidence on the value of reading aloud to and with children is now considerable. It is widely used and recommended with Polynesian children in New Zealand. It receives further support from the Fiafia evaluation described above, and is now being tried out in several Pacific Island countries.

In the past, abundant reading has been recommended as a means of acquiring information, as a way of gaining insight into human nature, as a method of enriching our leisure hours, and bringing us into contact with the minds of literary giants. The evidence is mounting that it is also a very effective means of learning language, of extending vocabulary, and of coming to grips with new syntax. There is reason to believe that bilingual children can benefit from reading in all these ways. It is surely no coincidence that the vast majority of competent bilinguals whom the writer has met in the South Pacific were voracious readers in their youth. They probably learned their English outside the classroom.

Let us not then cut our children off from a rich and valid source of learning by artificially pruning their language and presuming that we can control just how and in what order they will learn. Their natural language learning capacity is greater than we thought. We are just beginning to learn at what stages in their development, and with what kinds of materials and methods we can exploit it fully.

References

CHING, D.C. *Reading and the bilingual child.* Newark, Del.: International Reading Association, 1976.

CLAY, M.M. Language skills: A comparison of Maori, Samoan and Pakeha children. *New Zealand Journal of Educational Studies,* 1970, *5*(2), 153-162.

CLAY, M.M., GILL, M., GLYN, T., MCNAUGHTON, T., & SOLOMON, K. *Record of oral language.* Wellington, N.Z.: Educational Institute, 1976.

COOK, V.J. The analogy between first and second language learning. *International Review of Applied Linguistics* (IRAL), 1969, *7*(3), 207-216.

DE'ATH, P. *Interim guide notes for teaching, trialing and further development of the Fiafia reading program on Niue.* Unpublished notes, Niue Education Department, 1978.

DILLER, K.C. Some new trends for applied linguistics and foreign language teaching in the United States. *TESOL Quarterly,* 1975, *9*(1), 65-74.

DODSON, C.J. *Language teaching and the bilingual method.* London: Pitman, 1967.

DULAY, H.C., & BURT, M.K. Should we teach children syntax? *Language Learning,* 1973, *23*, 245-258.

ELLEY, W.B. *Learning to read in the South Pacific.* Paper presented at IRA Annual Convention, Atlanta, 1979.

ELLEY, W.B., & MANGUBHAI, F. Research project on reading in Fiji. *Fiji English Teachers Journal,* 1979, *15,* 1-7.

ELLEY, W.B., WATSON, J., & COWIE, C. *The impact of a book flood.* Wellington: New Zealand Council for Educational Research, 1976.

ERVIN-TRIPP, S. Structure and process in language acquisition. In J. Alatis (Ed.), *Twenty-first Annual Round Table.* Washington: Georgetown University Press, 1970.

GOODMAN, K.S. Miscue analysis: Theory and reality in reading. In D.B. Doake & B.O. Rourke (Eds.), *Proceedings of the 5th New Zealand Conference on Reading*, N.Z.E.I., 1976.

HANAYAN, E., MARKMAN, D.R., PELLETIER, S., & TUCKER, G.R. Differences in performance in elicited imitation between French monolingual and English speaking bilingual children. *International Review of Applied Linguistics*, 1978, *14*(4), 330-339.

KENNEDY, G.M. Conditions for language learning. In J.W. Oller & J.C. Richards (Eds.), *Focus on the learner*. Rowley, Mass.: Newbury House, 1973.

LAMBERT, W.E., & TUCKER, G.R. *Bilingual education of children: The St. Lambert experiment.* Rowley, Mass.: Newbury House, 1972.

LAOSA, L.M. Review of P.A. Hornby (Ed.), Bilingualism: Psychological, social and educational implications. *American Educational Research Journal*, 1978, *15*, 582.

MACNAMARA, J. *The cognitive strategies of language learning.* Paper presented at Conference on Child Language, Chicago, November 1971.

NALDER, S. *A global description of the book experience approach.* Unpublished notes, Auckland, N.Z., Education Department, 1975.

PITMAN, G.A. SPC/Tate English language programme and materials. *South Pacific Bulletin,* 1974, Second Quarter, 63-67.

PORTER, F. *The teaching of reading: Teaching children what?* Paper presented at Tonga Teachers College Seminar, September 1977.

RITCHIE, J. *Chance to be equal.* Cape Catley, Queen Charlotte Sound, N.Z., 1978.

RUBIN, J. *What the good language learner can teach us. TESOL Quarterly*, 1975, *9*(1), 41-52.

SHARP, D. *Language in bilingual communities.* London: Edward Arnold, 1973.

SMITH, F. *Reading.* New York: Cambridge University Press, 1978.

STAMP, D. et al. A readability survey of UNDP material using the cloze procedure. *Directions*, 1979, *2*, 23-28.

STREVENS, P. *New orientations in the teaching of English.* Fair Lawn, N.J.: University Press, 1977.

TATE, G.M. *Teaching structure: A teacher's handbook.* Wellington, N.Z.: A.H. & A.W. Reed, 1967.

TATE, G.M. *Oral English handbook.* Wellington, N.Z.: Reed Education, 1971.

TOWNSEND, H.E.R., & HOMER, D. *Gilbert Islands Project, June 1978.* South Australia: Salisbury College of Advanced Education, 1978.

TWADDELL, W.F. Vocabulary expansion in the TESOL classroom. *TESOL Quarterly,* 1973, *7*(1), 61-78.

WHEELER, J. *Reading to some purpose.* Paper presented at Language Teachers Workshops, Wellington, N.Z., May 1974.

Characteristics of Exemplary Reading Programs

S. Jay Samuels
University of Minnesota

Increasing Literacy Requirements and the Need for Quality Programs

In a technologically advanced society one of the important functions of the school is to educate its students so that they will be literate. A historical look at the concept of literacy reveals rather dramatic increases in the amount of reading skill required in order to be considered literate (Resnick & Resnick, 1977). With the inflation in literacy skills brought on in part by the ever increasing demands of our technologically oriented society, the quality of the school reading program has taken on added importance.

Because improving the quality of the school reading program is considered to be an important endeavor, educational researchers are currently attempting to specify the characteristics of successful reading programs with the underlying hope that by specifying the components which successful programs share in common, it will be possible to upgrade the quality of the less successful reading programs.

Purpose of This Article

Researchers are not the only ones who want to upgrade the quality of reading programs. Others who share this goal are school administrators, faculty, and the public. It seems as though schools are under increasing pressure to be accountable for the academic progress of their students. The increase in accountability is made manifest, in several ways, through local and statewide evaluations of academic progress, reports in local newspapers regarding the scholastic standing of each of the schools in the district, competency based promotion and graduation requirements, and law suits brought against school districts for what amounts to educational malpractice.

This article has been written as a response to the increased need to improve the quality of school reading programs. In pursuit of this goal, this article will describe the characteristics of successful and unsuccessful reading programs, and will touch upon the difficulties which school districts have encountered in trying to maintain quality reading programs.

The Need for Comprehensive Reading Programs

In this discussion of program quality, it should be pointed out that the term *program* has a broader meaning than the term *method*. In reading, we may refer to a phonic method, a look-say method, or a linguistic method. As viewed here, the reading method used by a teacher is merely a component within a larger entity called the reading program. Of course, the reading method one uses is important to overall program quality, but other components are important as well. For example, in addition to the reading method, other components which can affect program outcome include staff, administration, parents, and students. Thus, a program can be thought of as a set of interrelated components consisting of human, material, and procedural factors which influence the extent to which institutional objectives are realized. Armor, Osegura, Cox, King, McDonnell, Pascal, Pauly, and Zellman summarized their study of program characteristics and achievement by stating "...it is important to emphasize that no single school or classroom factor taken by itself is likely to produce large increases in reading achievement.... Reading instruction is far too complex to allow for simple policies or 'quick fixes'."

The plan which will be followed in this article is to describe briefly each of the program characteristics for a number of exemplary reading programs. Following the description of the programs, there will be a summary which will attempt to delineate the common characteristics across those exemplary programs described in this article. Finally, there will be a brief section describing characteristics of unsuccessful programs as well as the problems one encounters in trying to maintain program quality over time.

Components of Exemplary Reading Programs Found in Six Reports

Weber (1971)[1]. In this project, Weber looked for ghetto schools where reading scores were at the median for the nation as a whole. The four innercity schools in which Weber conducted his study were P.S. #11 and P.S. #129 in New York City; Woodland School in Kansas City, Missouri; and Ann Street School in Los Angeles, California. Since the students in these ghetto schools were from low socioeconomic level homes, one might have predicted that the level of academic achievement would be low. Instead, the level of achievement was approximately equivalent to the reading achievement found in the average income schools of the United States.

The four successful schools shared certain program characteristics. Weber identified these shared characteristics as:

1. Strong leadership. Administrators were aware of the problems and strongly supported efforts to raise achievement.
2. High expectations. Adminstrators, teachers, students, and parents were convinced that the students could succeed. Failure was unacceptable.
3. Both task and human relations orientation. The schools had sense of order and discipline. The mood was described as pleasant and happy. Although the teachers placed great emphasis on academic achievement, there was equal emphasis given to helping

---

[1] The criterion of success which Weber used was a national grade norm score as a median. In addition, a "successful" school had to meet another test: "that the percentage of gross failures be low." Weber (1971, p. 5) wrote, "typically, inner-city schools not only have a low achievement median, but the number of gross reading failures—children achieving far below national norm levels—is high." The third grade was selected as the grade to test for success since it was at this grade that the mechanics of reading should be mastered.

students feel good about themselves, their classrooms, and their racial origin.

4. Teacher aides. Where teacher aides were used, they were directly involved in instruction. Use of teacher aides in direct instruction changed the student-teacher ratio. Thus, instead of one teacher to thirty students, there might be two, three, or four teachers working with the students.
5. Additional reading personnel. The reading specialist worked directly with teachers and teacher aides, to upgrade teacher skills and to provide teachers with needed materials. This is known as the multiplier effect. If the specialist works with ten children, only ten children are helped; but if the ten teachers are helped, who in turn work with 30 children, then in effect, 300 children are helped.
6. Decoding emphasis. An early emphasis was given to what may be called a code-breaking subskill approach to reading rather than a holistic meaning emphasis. This should not be taken to mean that comprehension was unimportant, but that an early focus was given to those skills which facilitate word recognition.
7. Evaluation and quality control. There was continuous evaluation of student progress, and evaluation data served as a basis for diagnosis of student difficulty and the remediation of those problems.

Hawkridge, Tallmadge, and Larsen (1968)[2]. This survey is important because eighteen successful reading programs were matched against eighteen unsuccessful programs. The following differentiated the successful from the unsuccessful programs:

---

[2] This study matched successful and unsuccessful programs on the basis of age of students, ethnic composition of students, number of students served, and types of learning evaluated. The successful programs were drawn from an earlier study (Hawkridge, Chalupsky, & Roberts, 1968) in which 1,000 compensatory education programs were surveyed and 21 of these were selected because they had clear program descriptions, appropriate sample size, description of tests, evidence of reliability and statistical significance of the posttest. In addition, the program had to demonstrate "measured benefits," which meant that scores on standardized tests had to improve more than they would have under regular school programs.

1. Explicit goals. The program goals were clearly stated.
2. Focused instruction. Classroom instruction was aimed at achieving the objectives.
3. Relevant materials. The materials used were appropriate for the methods used and the goals.
4. Efficient use of time. There was high intensity of treatment.
5. Diagnosis. The learner's progress was monitored and checked at regular intervals and appropriate instruction given to each student to overcome individual problems.
6. Teacher training. Additional training was given to teachers before or during the program.
7. Parental involvement. Parents were utilized either as aides or to help students at home.

Wargo, Tallmadge, Michaels, Lipe, and Morris (1972)[3]. The purpose of this study was to obtain a detailed listing of exemplary reading programs, define the components of each of these programs, and then look for common elements across programs. The shared characteristics are listed below:

1. Academic objectives clearly stated and/or carefully planned.
2. Teacher training in the methods of the program.
3. Relevant instruction.
4. High treatment intensity.
5. Active parental involvement.
6. Personnel committed to the objectives and procedures.
7. Close supervision of teachers and aides.
8. Utilization of additional personnel in the form of teacher aides and/or reading specialists.
9. Continuous assessment system providing feedback and diagnostic information—immediate feedback of results.

---

[3] Twenty-one exemplary reading programs identified from earlier surveys were analyzed for characteristics shared in common.

10. Structured environment to aid student on task behavior.
11. Intensive staffing—use of aides.

New York State Office of Education (1974)[4]. The purpose of this study was to compare innercity elementary schools which were matched on social and cultural variables but which differed in reading achievement. In the successful schools, the administrator had a significant impact on the school's effectiveness. The administrator helped to develop plans for the reading program, provided planning time for the program, and helped to implement the program plan. In the unsuccessful school, the administrator took little direct interest in the reading program.

In the unsuccessful school, teachers tended to blame nonschool factors for the low achievement of the students, whereas, in the successful school, the responsibility and locus of control were placed within the school and under the control of teachers. Teachers in the unsuccessful school were pessimistic about their impact on students.

The successful school climate was one of high expectations, blended within a warm, humanistic, and rewarding atmosphere; whereas, in the less successful school, expectations were lower and there was less of a human relations orientation.

In summary, the successful school was marked by strong administrative leadership, positive expectations for success, extensive pupil evaluation, and special training of teachers.

CRAFT Project Harris and Serwer (1966)[5]. The purpose of this study was to compare two approaches to the teaching of reading and to determine the effect of time devoted to reading on the reading gains of students. After one year, the skills centered and the language experience centered approaches were compared, and a small but consistent difference favoring the skills centered over the language experience method was found. This finding is consistent with the emphasis on phonics, decoding, and subskills emphasis found in other programs.

---

[4] Two New York City elementary schools were studied to determine what specific school factors influenced reading achievement. Schools were matched for socioeconomic status race, and culture; but one school had significantly higher achievement in grades two, four, and six on three measures: functional reading level, word recognition, and comprehension.

[5] This study concerned itself with gains in reading achievement and whether those gains resulted from the *type* of approach used in reading or the *amount of time* spent teaching reading. A total of 1,141 disadvantaged pupils were drawn randomly from 12 schools located in the black ghetto areas of New York City. The teachers were volunteers within the system and were trained extensively in the specific approaches to be utilized. Both the skills centered approach and the language experience approach were compared. Pre- and posttests were administered to students with the posttest scores adjusted on entering scores to account for differences in entering skills.

Another finding of interest from the CRAFT Project dealt with the relationship between time allocated to reading and achievement. "Time allocated to reading" has two components, actual reading time and some residual, which represents factors such as handing out books, changing groups, and giving instructions. As may be expected, this aspect of the analysis showed that actual reading time was significantly correlated with achievement, and that when large periods of time allocated to reading were spent on nonreading activities, such as handing out books and giving directions, the results were unfavorable to achievement.

Direct Instruction Model (1977)[6]. The purpose of the study, which was called "Project Follow Through," was to compare each of nine different approaches on the learning of basic academic skills. The approaches, or models, varied from global whole-person, individualized child-centered, open classroom, language experience approaches, on the one hand, to skills emphasis, behaviorally oriented models. What is interesting to note is that the students in the Direct Instruction Model, which focused upon mastery of component skills in reading, and which utilized principles derived from the psychology of learning and instruction, ranked at the top, not only in subject matter learning but in measures of self-esteem, higher even than students who were in programs where building self-esteem was considered an important goal.

---

[6] Project Follow Through was viewed as a program which followed through with the educational efforts initiated with Head Start. The project was administered by the U.S. Office of Education, and eventually came to serve 75,000 low-income children annually under the guidance of what may be thought of as nine models of education. Of the nine models, the Direct Instruction Model was rated number one in reading. In the strictest sense, the project did not use an experimental design, but each model was compared to its own control. The comparisons for each model and its control were made on tests of academic skills, cognitive-conceptual skills, and measures relating to self-image. The basic data from the above tests were collected by Stanford Research Institute and analyzed by Abt Associates.

The Direct Instruction Model has the following characteristics:

Assumptions

1. All children, regardless of socioeconomic status, are capable of mastering basic academic skills.
2. Teaching failure is not excused.
3. Children from disadvantaged homes have less well developed prerequisite skills necessary for academic achievement.
4. Students from disadvantaged homes must be taught more in less time in order to catch up with other students.

The Rationale for Six Essential Teaching Components in the Direct Instruction Model

1. Teach the general case. By teaching a subset, the whole set is learned. For example, by teaching 40 sounds and skills for blending them together and saying them fast, generalized decoding skill, relevant to one-half of the common English words, is learned.
2. Use teacher aides in direct instruction. This increases the number of instructors in a classroom.
3. Daily program is carefully structured. Routines are established so that time is used efficiently.
4. Maintain student attention. This is done through rapid paced, teacher-directed, small group instruction.
5. Provide training and supervision. Teachers are carefully trained and supervised to ensure that appropriate skills are taught and utilized in the classroom.
6. Maintain quality control. Student progress is monitored with bi-weekly, criterion-referenced tests which help to detect problems while there is time to correct them.

The Rationale for Distinctive Features of the Direct Instruction Model

1. Scripted presentation of the lessons. The purpose is to provide quality control in the instructional delivery system.
2. Small group instruction. It is more efficient than one-to-one instruction and allows for better supervision than large group instruction.
3. Reinforcement. Although most people would agree that learning should be its own reward, there are many students who need extensive rewards to encourage learning. Hence, token economies may be used for students who do not respond to games, praise, or attention.
4. Training and supervision. Teachers are trained to use this approach to reading and are supervised to ensure implementation of instructional methods.
5. Program design. Utilizes task analysis, specifying objectives, analysis of objectives into component subskills, identifying prerequisite skills, selecting examples and sequencing the skills.

Wilder's (1977)[7] *Educational Testing Service Study of Exemplary Compensatory Reading Programs.* On the basis of achievement test scores, on-site visits and questionnaire data, five schools with exemplary compensatory reading programs were selected. What follows is a description of their characteristics.

---

[7] In 1971 the United States Office of Education funded this study of compensatory reading programs in grades two, four, and six. ETS conducted this study of characteristics of examplary reading programs with the premise that "the only real difference between compensatory and non-compensatory reading instruction lay with the source of the funding support, and consequently the study should be free of artificial restraints such as funding sources, materials, and characteristics of students. The definition of compensatory reading instruction which was adopted was any reading instruction provided for students because they were reading below grade level. 'Reading program' was broadly conceptualized to include the totality of a school's reading instruction including demographic and sociological features."

The effectiveness of the reading program was determined by a school's posttest reading scores in relation to its pretest scores: Schools which performed better than expected were selected for further study. By means of analysis of achievement scores and site visits an original sample of 741 schools was narrowed down to five exemplary schools which were examined in depth for their characteristics.

In all of the successful schools:

1. Reading was given top priority. This priority was made evident through time spent in reading activities and money spent on materials and resources.
2. There was effective educational leadership, either through the office of the principal or a resource person.
3. Attention to basic reading skills was the essence of the program.
4. A variety of materials was used. If a student encountered difficulty with a skill, alternative methods and materials were available to teach the skill.
5. There was discussion and cross-fertilization of ideas about reading. This took place as part of the inservice programs which are part of the development and maintenance of strong programs of reading.

Summary of Components of Successful Reading Programs across Investigations

Having described the components of the successful reading programs for each of the investigations, it is possible now to examine the extent of overlap which these programs share. Table 1 provides a convenient means for visual inspection of the program characteristics found in each of these successful approaches to reading instruction.

As mentioned previously, a good program is considerably more than a reading method. It may include factors and variables as subtle as assumptions about the ability of children to master basic reading skills, or the responsibility which the school has for helping children master basic academic skills. In addition to the assumptions which underlie the programs, other important factors include personnel, training, application of principles of learning, and evaluation.

At the possible risk of overgeneralization, how might one summarize the components found in successful reading programs?

Table 1

Components	Weber (1971)	Hawkridge, Tallmadge, and Larsen (1968)	Wargo et al. (1972)	CRAFT (1966)	NY State Office of Ed. (1974)	Direct Instruction (1977)	Wilder (1977)
Strong administrative leadership	+	+	+		+	+	+
High expectations and belief that the teacher can make the difference	+		+	+	+	+	
Teacher aids used in direct instruction	+	+	+			+	
Reading specialist	+		+			+	+
Teacher training	+	+	+		+	+	+
Teacher role — present, diagnose, remediate	+	+	+			+	
Specific reading objectives	+	+	+	+		+	+
Skills centered curriculum	+		+	+		+	+
Instruction and materials relevant to goals	+	+	+			+	+
Structural environment	+		+		+	+	
Positive classroom climate	+		+		+		
Efficient use of time	+	+	+	+		+	+
High intensity of treatment	+	+	+	+		+	+
Frequent evaluation of student progress	+	+	+		+	+	
Supervision of teachers			+			+	

A + means that component was identified in the report.

1. *Underlying Assumptions*

There are several important assumptions or beliefs held by staff in exemplary reading programs which seem to be associated with student academic growth. The first assumption one often finds is that the school can have a significant impact on the academic achievement of its students. Consequently, the school is responsible to a large extent for both the success and failure of its students. The second assumption is that most children are capable of mastering the basic academic skills. If one assumes these two premises, namely that the school is responsible for student achievement and that most students can master basic academic skills, then one does not routinely absolve the school from the responsibility for student failure with explanations such as the student had "poor motivation," "lack of readiness," or "inadequate home background conditions." Instead, when students fail to learn, the school assumes the major responsibility for identifying and remediating the problem.

2. *Personnel*

Administrators. One frequently finds a strong administrative leader associated with the exemplary reading program. In fact, the Rand Corporation Report (Berman & McLaughlin, 1978) on implementing and sustaining educational innovation states that the importance of administrators to the creation and maintenance of innovative practice can hardly be overstated. Administrative support, in the form of encouragement and the creation of an organizational climate which gives the project "legitimacy" is essential to the project's development, operation, and maintenance. The support which the administrator gives is not in "how to do it," but in providing time for planning and carrying out decisions, securing necessary financial support, and running interference against any counterforces in the school or community.

Teacher aides. In order to reduce the student-to-teacher ratio, teacher aides were often used in direct instruction in the exemplary projects. The aides had received training, so that an observer in a classroom might find what appeared to be anywhere from two to four teachers in a classroom. In actuality, what one was observing was a teacher and several trained teacher aides working directly with students.

Reading specialists. A number of projects used specialists in a variety of ways, from helping to train staff, to providing materials, to serving as on-the-job consultants. The specialists did not work as remedial education teachers but worked directly with teachers and teacher aides.

Teachers. The belief system of the teacher is important. One important element in the belief system is a feeling of commitment, dedication, and support for the project goals. The feeling of commitment and support becomes the energizing force which motivates the teacher to provide the time and energy necessary to implement the project goals. The teacher who believes that student success or failure depends upon what happens in the classroom and that student failure is not acceptable tends to be found in the exemplary reading programs. In the Rand Report, Berman and McLaughlin (1978) state that teachers rise to challenges. Ambitious and demanding projects seem more likely to win the commitment of teachers than routine projects because the demanding project appeals to a teacher's sense of professionalism. A major reason for teachers taking on the extra work and disruption of attempting to bring about change in themselves and their students is the belief that they will become better teachers and the students will benefit.

3. *Teacher Training and Supervision*

There seems to be strong support across investigations of exemplary reading programs for teacher training. Research by Berman and McLaughlin (1978) indicates that educators and goverment officials have been too optimistic about how much time is needed to produce change in an educational system. Their research indicates it takes two years to get a project off the ground, two years to implement a project, and an additional two years to produce a stable effect on student achievement scores. Thus, it might take six to seven years before one could expect stable achievement gains.

The teacher training which takes place occurs during the two year start-up time and the two year implementation time.

The Rand Report states that helpful training experiences were:

Concrete, teacher-specific
Regular meetings which focused on practical problems
Teacher participation in decisions
Teacher observation of similar projects or model lessons

Additional factors found in the successful innercity schools were:

Classroom assistance from project staff or reading specialists
Local materials developed
Principal participation in training

In order to insure implementation of prescribed teacher practices, some of the projects supervised instruction.

4. *Curriculum*

Objectives. The successful reading projects had objectives which were clear and specific. Although the primary focus was what might be thought of as "task orientation," there was a "human relations orientation" as well. The twin goals reflect the fact that the best environment in which to help master academic skills is one in which students find the classroom to be a friendly, warm, and supportive place to work. In other words, when there is task orientation combined with positive classroom climate, the situation is conducive to academic growth.

Skills centered curriculum. This approach to reading utilized what may be called a "subskill" orientation to reading. With this approach fluent reading is thought of as a highly complex behavior which can be analyzed and separated into its components through a process called task analysis. The components are then taught to beginning readers in a sequence which is determined partly by the task analysis, partly by logical analysis, and partly from what we have learned through research on instruction. In order to help the student integrate the subskills into the higher-order behavior required in skilled reading, in addition to the instruction in subskills, the student is given ample practice in reading meaningful and interesting material in context.

273

Relevant instruction and materials. The successful reading programs had clear and specific objectives and the subskills which were taught were derived by means of task analysis. What one tends to find in the exemplary programs is that both the type of instruction and the materials were appropriate and relevant for achieving the specific objectives.

5. *Time*

Time is an important factor to consider in learning and achievement. While it is essential to allocate a sufficient amount of time so that learning can take place, it is equally important that the available time is used efficiently. In successful programs, teachers devoted more class time to task-related activities, and the major part of the school day was spent in structured activities that left little unoccupied time. In addition, instruction was kept at a low level of complexity. The classrooms were described as being orderly, since less time was wasted on discipline problems and giving instruction on routine matters such as passing out books and transition from one activity to another. In summary, in several of the successful programs, extra amounts of time were allocated to reading; and in all programs, time was used efficiently.

6. *Structure*

Stucture is an important vehicle for utilizing time efficiently. In a sense, structure is part of good management skills in which routines are established in order to facilitate the movement of students and materials. The successful programs helped teachers to establish good management skills so that time would not be wasted. This concept can be extended to the point where the most efficient ways to communicate ideas to students were developed and shared with teachers.

7. *Evaluation*

Student progress is monitored through frequent testing on each of the subskills so that quality control can be maintained through diagnosis and remediation of student problems.

Characteristics of Unsuccessful Reading Programs

While it is useful to know the characteristics associated with successful reading programs, if we wish to build programs which will increase reading achievement, it is probably equally important to know what is associated with those programs which were unsuccessful. With this end in mind, we shall summarize several reports which have described characteristics of programs which did not significantly improve achievement (Armor et al., 1976; Berman & McLaughlin, 1978).

1. *Improper program initiation and implementation.* Before a program can work there must be a climate of enthusiasm, commitment, and support for it from the personnel. It is this climate of commitment and support which motivates the staff to implement and translate the project goals into effective classroom practices. The following approaches to initiation fail to create the necessary climate for support and implementation.

Opportunism. The motivation for the initiation of the project is opportunistic, representing a response to political pressure or an attempt to get federal money. Opportunistic initiation is characterized by lack of support and commitment from both central office and project staff. Since there is lack of support at any level, the staff is unwilling to invest the resources, time, and energy necessary to implement the project.

Top-down motivation. Although the central office staff is sincerely interested in the program, they fail to win the support of the project staff. Consequently, the project staff does not invest the time and energy necessary to implement the project.

Bottom-up motivation. The "grass roots" enthusiasm of the instructional staff is not matched by the central office staff. Consequently, there may be inadequate project implementation because the central office fails to allocate a sufficient amount of resources and staff services.

2. *Inadequate incubation and development time.* A reading program is broad in scope, incorporating elements which include personnel, methods, materials, finances, and clients. Changes which affect this many elements require considerable

periods of time for planning and implementation. Current research suggests that significant innovation requires two years of planning and incubation time, two years for implementation, and two years to produce a stable effect on student achievement. Those projects which were unsuccessful rushed forward too soon and often failed to create the necessary climate for district support and commitment.

3. *Narrow, piecemeal approaches.* There are so many important elements in a good reading program, that simple, "quick fix," single element approaches usually cannot produce a significant impact on achievement. To produce a significant impact, a comprehensive approach operating on a student over time is required.

4. *Lack of commitment, feelings of inadequacy.* The less successful projects tended to have staff which either did not have a feeling of commitment and dedication to the project plan or lacked a sense of efficacy regarding their ability to bring about the necessary changes.

5. *Unsupportive principal.* The school principal is a vital element, and in the unsuccessful projects, the principal often failed to create norms which emphasized hard work and dedication and an environment which supports teacher efforts and protects the teacher from disruptive forces.

6. *Inefficient use of time.* Time is a critical variable in learning. It is obvious that if mastery is to occur, the amount of time allocated must be commensurate with the needs of the students. However it is important that the allocated time be used efficiently. In the unsuccessful classrooms, time was wasted because routines were not established and there were often interruptions brought about by discipline problems.

Problems in Maintaining Quality Programs

The purpose of this article has been to delineate the components found in successful and unsuccessful reading programs in the hope that this information would be of help to schools in establishing more effective environments in which students would acquire literacy skills. Although the information regarding components of successful programs is now available, this information is only a starting point.

Schools which currently have successful programs are faced with the task of maintaining their quality programs, while the less successful schools have additional obstacles to overcome. They first must improve their current programs and, once they have improved their programs, they must then endeavor to maintain them at quality levels. The Rand Corporation Report (Berman & McLaughlin, 1978) contains some sobering conclusions regarding the establishment and maintenance of quality programs. First, they found that successful projects were not easily disseminated to new sites and the performance levels in the new sites often fell short of the performance level in the original site. Second, even the original successful projects had difficulty sustaining their success over a number of years. Third, often when the original funding agency withdrew its financial support, the school district was unprepared to provide the level of support necessary to sustain the project.

These sobering findings regarding the difficulty of fostering and maintaining change are neither surprising nor unique to the field of education. We still have much work to do. While we seem to know a considerable amount about the characteristics of successful reading programs, we still need to know more about which of these characteristics uniquely define the successful programs and are not present in the unsuccessful programs. Although the goals are difficult to achieve, the importance of the task is such that it seems to be well worth the effort.

References

ARMOR, D., OSEGURA, P., COX, M., KING, N., MCDONNELL, L., PASCAL, A., PAULY, E., & ZELLMAN, G. *Analysis of the school preferred reading program in selected Los Angeles minority schools.* Los Angeles: Rand Corp., 1976.

BECKER, W.C. Teaching reading and language to the disadvantaged—what have we learned from field research? *Harvard Educational Review*, 1977, *47*, 518-543.

BERMAN, P., & MCLAUGHLIN, M.W. *Federal programs supporting educational change, Vol. VIII: Implementing and sustaining innovations* (USOE). Santa Monica, Calif.: Rand Corp., 1978.

HARRIS, A.J., & SERWER, B.L. The CRAFT Project: Instructional time in reading research. *Reading Research Quarterly*, 1966, *2*(1), 27-57.

HAWKRIDGE, D.G., CHALUPSKY, A.B., & ROBERTS, A.O. *A study of selected exemplary programs for the education of disadvantaged children.* Palo Alto, Calif.: American Institutes for Research, 1968. (ERIC ED 023-776, 023-777)

HAWKRIDGE, D.G., TALLMADGE, G.K., & LARSEN. *A study of selected exemplary programs for the education of disadvantaged children* (Final Reports 1 & 2). Palo Alto, Calif.: American Institutes for Research, 1968. (ERIC ED U23776-ED U23777)

NEW YORK STATE OFFICE OF EDUCATION PERFORMANCE REVIEW. *School factors influencing reading achievement: A performance review.* Albany, N.Y., March 1974. (ERIC ED 089 211)

RESNICK, D.P., & RESNICK, L.B. The nature of literacy. *Harvard Educational Review*, 1977, *47*, 370-385.

WARGO, M.J., TALLMADGE, G.K., MICHAELS, D.D., LIPE, D., & MORRIS, S.J. *ESEA Title I: A reanalysis and synthesis of evaluation data from fiscal year 1965-1970.* Palo Alto, Calif.: American Institutes for Research, 1972.

WEBER, G. *Inner-city can be taught to read: 4 successful schools.* Washington: Council for Basic Education, 1971.

WILDER, G. Five exemplary reading programs. In J.T. Guthrie (Ed.), *Cognition, curriculum, and comprehension.* Newark, Del.: International Reading Association, 1977.

Overcoming Educational Disadvantagedness

Harry Singer
Irving H. Balow
University of California at Riverside

Twenty-five years ago, the U.S. Supreme Court reversed its 1896 decision and ruled that "separate but equal" schools for blacks and whites were inherently unequal in educational opportunity and, therefore, unconstitutional. This decision initiated desegregation of school districts and introduced a new and different assumption into the concept of equality of educational opportunity, "the assumption that equality of opportunity depends in some fashion upon the effects of schooling" (Coleman, 1968, p. 15). Under the 1896 decision, judgments on equality of educational opportunity were based on equality of educational *input*, which was assessed by such criteria as cost of buildings and supplies, teachers' salaries and academic preparation, class sizes, etc. However, education of exceptional children, even prior to 1954, required *unequal input*: class sizes were much smaller and frequently required special facilities and equipment. In effect, the 1954 Supreme Court decision expanded the principle under which exceptional children had been educated to now include all ethnic groups.

Moreover, the responsibility for attainment of achievement shifted from the student to the educational institution. After tracing the evolution of the concept of equality of educational opportunity, Coleman (1968, p. 22), summarized the shift of responsibility in these words:

> The implication of the most recent concept...is that the responsibility to create achievement lies with the educational institution, not the child. The difference in achievement at grade 12 between the average Negro and the average white is, in effect, the degree of inequality of opportunity, and the reduction of that inequality is a responsibility of the school. This shift in responsibility follows logically from the change in the concept of equality of educational opportunity from school resource inputs to effects of schooling. When that change occurred, as it has in the past few years, the school's responsibility shifted from increasing and distributing equally *its* "quality" to increasing the quality of its *students'* achievements. This is a notable shift, and one which should have strong consequences for the practice of education in future years.

Thus, the Supreme Court decision had two major educational consequences:

1. Schools had to desegregate their students, whether or not they had been segregated a) *de jure*, for example, through gerrymandering of school boundary lines or b) *de facto*, for example, through residential location. Consequently, desegregation meant that schools had to seek mixtures of majority and minority students and eventually teachers in each school in proportions that approximated their representation in the population.

2. The concept of equality of educational opportunity had to be assessed on the effects of schooling. Of course, the concept of equal effects did not mean that all students were to be alike in achievement, but at least averages and ranges of individual differences in achievement among schools were to be approximately the same. Consequently, when a complaint was filed in court that a California school district was practicing *de jure* segregation because the reading achievement scores of students in its schools with predominantly black students were lower than those of students in its schools with a high percentage of white students, it was upheld by the court. Therefore, this school district was ordered to desegregate its students (Pena vs. Superior Court of Kern County, 1975).

But desegregation did not free a school district from the responsibility of attaining equality of output for its ethnic minorities. Compensatory education was still necessary because the magnitude of achievement among ethnic groups throughout the country was unequal, as shown by "Equality of Educational Op-

portunity" (Coleman et al., 1966), a survey more popularly known as the "Coleman Report."

Purpose of This Review

We shall review the Coleman Report and selected studies of compensatory education that have been conducted over the past 25 years. These studies include 1) the effects of desegregation upon the educational achievement of majority and minority groups in one school district and 2) the results of two major experiments to improve the education of disadvantaged children: "Project Head Start" and its sequel "Project Follow Through." Our purpose in reviewing these studies is to determine what effect the new concept of equality of educational opportunity has had upon the academic achievement of educationally disadvantaged children. We shall start our review with the Equality of Educational Opportunity Survey.

Equality of Educational Opportunity Survey

The Equality of Educational Opportunity Survey was authorized by the Civil Rights Act of 1964. Its purpose, as stated in Section 402 (Coleman et al., 1966, p. iii) was to "conduct a survey...concerning the lack of availability of equal educational opportunities for individuals by reason of race, color, religion, or national origin in public educational institutions at all levels in the United States, its territories and possessions, and the District of Columbia." The groups investigated were "Negroes, American Indians, Oriental Americans, Puerto Ricans living in the continental United States, Mexican Americans, and whites other than Mexican Americans and Puerto Ricans often called 'majority' or simply 'white'." The survey focused on four major questions. These are the Survey's questions and its answers, almost verbatim:

1. *What is the extent to which racial and ethnic groups are segregated from one another?* The survey found that the "great majority of American children attend schools that are largely segregated." Among minority groups, Negroes were most segre-

gated, that is, they attended schools where their fellow students were of the same racial background as they were. In fact, more than 65 percent of all Negro students were attending schools that were between 90 to 100 percent Negro. This figure changed by grade 12 to 66 percent who were attending schools that were 50 or more percent Negro. The same pattern tended to hold also for teachers of Negro and white, although not as strongly. "On a nationwide basis where teachers and students are not matched, the trend is in one direction: white teachers teach Negro children but Negro teachers seldom teach white children."

2. *Do schools offer equal educational opportunities in terms of input characteristics?* White children were attending elementary schools with a smaller average number of children per room (29) than any of the minorities whose class sizes ranged from 30 to 33, except in the non-metropolitan North and West and Southwest where the pattern was reversed.[1] At the high school level, the average was 22 white students per teacher and 26 Negro students per teacher; but in some regional areas such as the metropolitan Midwest, the difference was far greater: the average white had a ratio of 1 teacher to 33 students and the average Negro had 1 teacher to every 54 pupils per room!

In access to facilities, the Negro had fewer "facilities that seem most related to academic achievement: they had less access to physics, chemistry, and language laboratories, fewer books per pupil in their school libraries, and less sufficient supplies of textbooks. They also had less access to curricular (college prep) and extracurricular (debating and school newspaper) programs that would appear to have a relationship to academic achievement." Moreover, the average Negro pupil tended to have less able teachers than did white pupils as assessed by types of colleges attended, years of teaching experience, salary, and score on a 30-word vocabulary test. The classmates or peer group of the Negroes as compared with the whites were less often enrolled in college prep curriculum and had taken fewer courses in English, mathematics, foreign language, and science.

---

[1] However, the West, for example, California, was not free of *de jure* and *de facto* segregation and other educational policies and practices that were discriminatory against its minority groups (Hendrick, 1975).

3. *How much do students learn as measured by performance on standardized achievement tests?*[2] The Survey found that "with some exceptions—notably Oriental Americans—average minority pupils score distinctly lower on these tests at every level that the average white pupil." A difference was already present at the first grade level. On a standard scale where the average score is 50 and the standard deviation is 10 (which means that 16 percent are above 60 and 16 percent are below 40), the scores for Negro and white revealed these differences:

Tests	Median Scores	
	Negro	White
Nonverbal	43.4	54.1
Verbal	45.4	53.2

At twelfth grade, the differences had increased:

Tests	Median Scores	
	Negro	White
Nonverbal	40.9	52.0
Verbal	40.9	52.1

Tests	Median Scores	
	Negro	White
Reading	42.2	51.9
Mathematics	41.8	51.8
General Information	40.6	52.2
Average of the five tests	41.1	52.0

---

[2] The EEOS Survey used verbal ability as an index of achievement. Some evaluators agreed with this use of a verbal ability test. For example, Dyer (1972, p. 516) stated that a verbal ability test "correlates so highly with measures of achievement in reading, mathematics, and factual information that it serves as a useful surrogate for the measurement of other forms of learning." However, verbal ability is also an index of general mental ability. As such, it is an index of both school *and* extraschool factors. Hence, it is not a *specific* nor a *sensitive* measure of school achievement.

In grade level equivalents, the gap between Negroes and whites showed a progressive increase with age and grade in school. In the metropolitan Northeast, the gap was about 1.6 years behind in grade 6, 2.4 in grade 9, and 3.3 in grade 12. However, some minority children, of course, performed better than many white children. For example, about 50 percent of the Negroes scored better than 16 percent of the whites.

4. *What is the relationship between achievement and school characteristics?* a) When socioeconomic factors were taken into account, differences between schools accounted for only a small fraction of differences in pupil achievement. b) Variations in facilities and curriculums of the schools accounted for relatively *little variation* in pupil achievement or standardized tests. However, variations made more of a difference for minorities than they did for whites; for example, existence of science laboratories showed a small but consistent relationship to achievement. c) Quality of teachers showed a progressively stronger relationship at higher grades; more so to minority than to majority students. The greatest relationship was first between teacher's score on the verbal test and student achievement and second between the teacher's educational level and student achievement. d) Pupil achievement was also "strongly related to the educational backgrounds and aspirations of other students in the school." The principal indices for these variables related to achievement were proportion of homes having encyclopedias and proportion of students planning to go to college.

After analyzing achievement scores of children of a given family background who had been put in schools of different social composition where the achievement was at quite different levels, the survey arrived at this conclusion:

> Thus, if a white pupil from a home that is strongly and effectively supportive of education is put in a school where most pupils do not come from such homes, his achievement will be little different than if he were in a school composed of others like himself. But if a minority pupil from a home without much educational strength is put with schoolmates with strong educational background, his achievement is likely to increase. . . .
>
> The average white student's achievement seems to be less affected by the strength or weaknesses of his school's facilities, curriculums, and teach-

> ers than is the average minority pupils. ...the inference might then be made that improving the school of a minority pupil may increase his achievement more than would improving the school of a white child increase his.

A mechanism that could mediate such an effect upon achievement of minority pupils in integrated schools is lateral transmission of peer group values (Wilson, 1963). This mechanism implies that minority pupils in integrated schools are influenced through interactions with classmates from the majority group and would consequently tend to acquire and act upon the values which underlie the achievement of majority pupils.

Although the inferences of the Survey and the mechanism of lateral transmission of peer group values appear to be plausible, the inferences of the Survey were based upon symptomatic, not upon experimental evidence. However, about the same time that the Coleman Report was formulating its inferences, an event took place in Riverside, California that provided a natural-type situation for experimentally testing some of them.

A Naturalistic Experiment in Riverside, California

In 1965 the Board of Education of the Riverside Unified School District acting unusually rapidly decided to desegregate its schools. Within seven weeks, from September 3 to October 25, 1965, the Board went from "initial minority protest" to "approval of a detailed integration plan" (Singer & Hendrick, 1967, p. 145).

The Riverside Unified School District is a medium sized school system with a total school population of 25,600 and a minority enrollment at the time of integration of 6.1 percent Black, 10.7 percent Mexican American and 1.7 percent other minority groups. Integration in Riverside consisted of closing down two of its three completely segregated schools and phasing out the third. Pupils from the segregated schools were then bussed to "receiving schools." The school board's plan called for minority enrollment in each school to approximate the same percentage as was enrolled in the district. With this decision on October 25, 1965, "*de facto segregation* had been changed by virtue of board policy into *de facto integration*" (Singer & Hendrick, 1967, p. 145).

Although this integration policy was justified on the basis of moral, social, educational, and legal reasons, including a broad interpretation of the 1954 Supreme Court mandate (Hendrick, 1968), school district personnel and university professors joined in a cooperative venture to evaluate the effects of integration on achievement and adjustment of the Anglo, Black, and Mexican American groups. However, this report focuses only on the reading achievement aspect of this longitudinal investigation, including on only those students who were tested after integration had occurred and who were then retested in grade 3.9 and again in grade 6.2.[1] Only those students who were tested at each of these times were included in the "purified" longitudinal sample. Students who were not tested at each time or who had moved away were not included. The purified sample was then compared with a cross sectional sample which had been tested in 1965 and constituted a "baseline" control group. Although it would have been more desirable to have had a longitudinal control group that did not experience segregation, the 1965 cross sectional group nevertheless appears to have been an adequate control group as indicated by the results of the experiment. Both the longitudinal and the baseline control groups took the same Stanford Achievement Tests (SAT) in grades 1.9 and 3.9, but in grade 6.2 the cross sectional group took the SAT while the longitudinal group had the Comprehensive Test of Basic Skills. The tests that were administered are shown in Table 1.

Table 1
Tests Administered to Longitudinal and Cross Sectional Groups

Grade	Test
1.9	Stanford Achievement Test, Primary I, Form W
3.9	Stanford Achievement Test, Primary II, Form X
6.2	Stanford Achievement Test, Intermediate II, Form W, (cross sectional group) and Comprehensive Test of Basic Skills (longitudinal group).

---

[1] Much of the information reviewed here has already been reported in Singer (1970) and in Singer, Gerard, and Redfearn (1975).

The grade equivalent scores for the longitudinal and cross sectional groups in grades 1 and 3 were based on the same test and norm groups, and therefore are directly comparable. But the longitudinal and cross sectional group took different tests in grade 6; therefore, the grade equivalent scores for these groups are only comparable to the extent that the tests are highly correlated and the norm groups are equivalent.

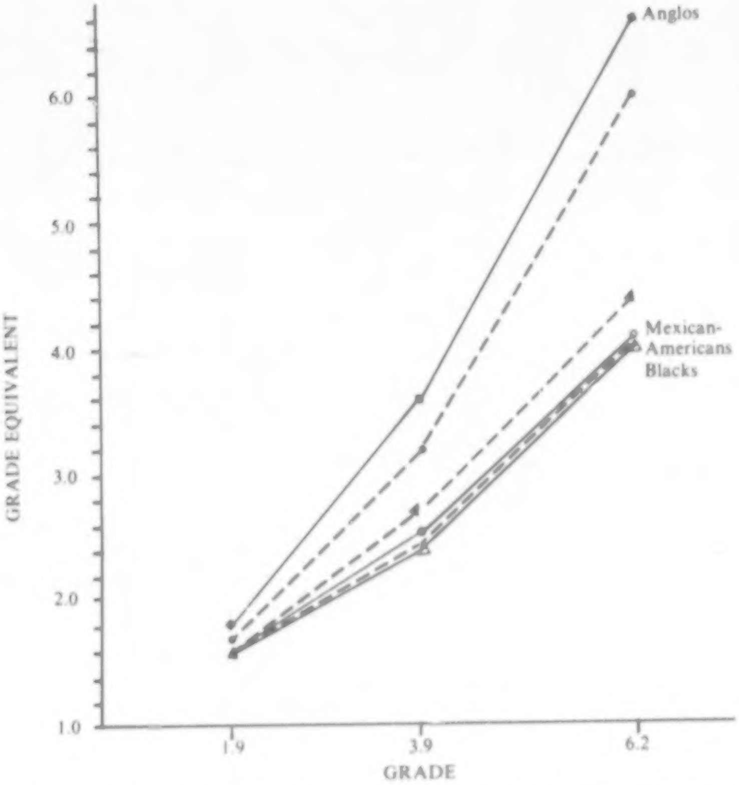


Figure 1 Mean reading achievement for the longitudinal (solid lines) and cross sectional (dashed) samples of Anglos, Mexican-Americans, and Blacks.[4]

[4] Singer, H., Gerard, H., & Redfearn, D. Achievement. In H. Gerard & N. Miller (Eds.), *School desegregation: A long term-study*. New York: Plenum Press, 1975. Reprinted with permission.

The results of the reading test scores are shown in Figure 1. The differences at the third and sixth grade levels between the longitudinal groups and their respective baseline controls are not statistically significant. Instead, the trends for the longitudinal sample and their respective baseline control groups resemble each other so closely that they tend to justify the use of data on the 1966 groups as baseline control data. In general, these data indicate that integration had neither favorable nor deleterious effects upon the rates of development of the reading achievement of the majority and minority groups. The Anglos did not lose in achievement, as their parents had feared they would, nor did the minority groups gain in achievement, as their parents had hoped they would. The Anglos, Blacks, and Mexican Americans had grade level achievements in grade 6 that approximated the findings of the *EEOS* Survey: The mean Anglo reading achievement was at or slightly ahead of the grade equivalent norm group while the Blacks and Mexican Americans were about two grade equivalent levels below the norm. Thus, the natural-type experiment in the Riverside desegregation study did not support the inferences of the Coleman Report . This lack of support occurred despite a situation that should have been highly favorable to the Coleman Report's inferences since a) normalization of instruction, the tendency to teach towards the average level of the class, would have stimulated and perhaps motivated minority students towards higher achievement and b) lateral transmission of peer group values was more likely to occur in Riverside than in other places in the country where the percent minority is even higher, for even after integration the percent of majority students in Riverside constituted 80 percent of the class.[5]

---

[5] However, policies of reducing school segregation were "almost wholly ineffective in the largest school districts" in the United States. Moreover, in large cities, at least during the years 1968-1972, "whites with children in public schools are leaving cities with high proportions of blacks." Although these findings are tentative, they nevertheless led Coleman (1976, pp. 318-322) to suggest that the policy of desegregation in large school districts was not achieving its purpose of "increasing overall contact among races in schools." However, in the next to largest school districts, changes in white public school enrollment were not related to "the extent of desegregation nor changes in degree of racial proximity" (Jackson, 1976, p. 332). Hence, it appears that desegregation policy was having a differential effect related to district size: an opposite to intended effect in the largest school districts and its intended effect in smaller school districts.

Of course, when results of an experiment are negative, we can speculate on the possible reasons. One of these is that pupils should already have had the necessary motivational systems of values, attitudes, and beliefs as well as background experiences for transforming capabilities into achievement prior to schooling in order to attain maximal benefit from their education (Katz, 1968). Since reading readiness scores at the end of kindergarten, two years after integration had begun in Riverside, were still significantly different for Anglos vs. Blacks or Mexican Americans (Purl & Kleinke, 1968), ameliorative action needed to be taken at the kindergarten or at the preschool level to reduce this readiness differential, which is cumulatively amplified throughout the grades. Consequently, we shall now review some intervention experiments at the preschool level and ascertain their effects upon subsequent school achievement.

Preschool Experiments

In contrast to nursery schools and many kindergartens which at least until the 1960s emphasized adjustment, Bereiter and Engelmann (1966) organized a *preschool*, which they defined as a downward extension of elementary school. Using a task analysis approach to curriculum construction, they defined the criteria for readiness at the first grade level, and used them to formulate 15 specific, academic, minimum goals. They then selected content for a curriculum which would enable children to achieve these goals, even educationally disadvantaged children who had necessary mental capability but lacked educational stimulation for reaching their potential. Specifically, they found that children at the beginning of first grade are expected to answer simple questions with complete sentences, transform sentences from a positive to a negative statement, infer and state some polar opposites (large-little, hot-cold, day-night, etc.), use certain prepositions correctly, make some simple logical deductions, identify basic colors, count aloud to 20, count objects up to 10, name all the vowels and most of the consonants, rhyme words, and have a sight-reading vocabulary of at least four words. They then taught their students this content.

Pines (1967) observed Bereiter and Engelmann's preschool in action. She reported that the children, in groups of five, were taught in the most direct manner possible. For example, in language development, the children were taught as though they were learning a foreign language: the child repeated verbatim the teacher's initial statement about an object or a picture and then was expected to give a complete answer to a simple question about the object or picture. The program consisted of three instructional periods of 20 minutes each, separated by half hour recess periods for singing and games. Even during these recess periods, the children sang specially composed songs and played games that were designed to teach them the content that had been stressed during the instructional periods.

Although Bereiter and Engelmann's emphasis upon academic content at the preschool level was controversial for its time, Pines observed that their students seemed to enjoy the program and appeared to be making excellent progress toward their objectives. Her observations were correct: Bereiter and Engelmann reported that a group of 15 educationally disadvantaged children who had started in their preschool at a median age of 4 years, 6 months had achieved enough after 9 months of instruction to qualify in reading and arithmetic for entrance to first grade.

A similar type of preschool at the University of Chicago (Fowler, 1965) stressed general cognitive development. Children engaged in what appeared to them to be a game consisting of finding and manipulating objects, but the game was actually designed to develop cognitive abilities, such as making discriminations among objects and then abstracting and generalizing certain characteristics of these objects. For example, the children were instructed to find all the pictures which show people working in factories. For the development of language ability, the children participated in the formulation of stories narrated around objects and scenes manipulated by a storyteller. Since all objects were labeled and defined for the children, they also had an opportunity to develop their word perception along with their vocabulary.

The Bereiter and Engelmann program can be categorized as *"direct instruction,"* teaching students what they need to learn. In contrast, the Chicago preschool program fits into the rubric of indirect instruction emphasizing *cognitive development,* which provides instruction aimed at improving cognitive processes and abilities. This training was presumably necessary for subsequent academic achievement. Both types of programs were subsequently given field trials and compared along with other models in a social and experimental educational program known as "Follow Through," which was one of the two largest federal experiments for exploring alternative approaches for delivering educational and social services. The other, which preceded Follow Through and which had broader objectives than the preschool, but was also for disadvantaged children, was known as "Project Head Start."

Project Head Start

Even while desegregation was beginning as the sole compensatory education approach, later to prove inadequate for attaining the goal of equality of output, intervention programs were being initiated. The largest was the federally sponsored Project Head Start which began in 1965. It had the purpose of providing compensatory preschool education for poor children so that children from low income families could enter school on an equal basis with their more economically fortunate peers. The hope was that this type of intervention at the preschool level would contribute not only toward equality of educational opportunity but also toward the disruption of the cycle of poverty. The objectives of the program were therefore comprehensive; they included provisions for the child, the child's family, and relationships among the child, the family, and society. The provisions for the child were concerned with his or her health and physical abilities, mental processes and skills, expectations of success and confidence in future learning, relations with his or her family, attitude of responsibility toward society, and sense of dignity and self worth. The program also aimed to develop the ability of the child's family to relate in a positive way toward the child and to increase the family's sense of dignity and self worth. Lastly, the program was to provide opportunities for society to work with the poor on solving their problems.[6]

The comprehensiveness of Project Head Start's objectives had fostered the development of a variety of preschool models. In 1969, the U.S. Office of Education initiated an evaluation of the models that had been well formulated and tested, and whose sponsors[7] were using the principles of their models for implementing programs for elementary children in Project Follow Through. The purpose of the evaluation, which was known as the Head Start Planned Variation Study, was to provide information to communities on a) processes for implementing preschools, b) kinds of experiences provided in preschools, c) effects of different intervention programs on children and their families, d) comparison of intervention effects in preschool vs. primary grades, and e) the advantage of continuing the same preschool strategy into the primary grades.

Eight models met the criteria for inclusion in the evaluation. Table 2 contains brief descriptions of these models, which were grouped into three categories on the basis of their stated objectives and procedures. These categories, determined by the Stanford Research Institute which did the evaluation, were defined by Bissell (1973, pp. 69-70) as 1) *Preacademic*, which emphasizes development of preschool skills, such as number and letter recognition, reading, writing, and instructional language, and uses systematic reinforcement as an instructional technique; 2) *Cognitive Discovery*, which promotes growth in basic cognitive processes, such as categorizing, differentiating, abstracting, and inferring, by use of adult verbal accompaniment to children's sequenced exploration; and 3) *Discovery*, which encourages free exploration and self-expression, and stresses the "child's sense of self-worth, trust in adults and the world, and respect for others."

---

[6] A report on this project was made by Bissell (1973). Much of the information for this section comes from her report.

[7] A sponsor is anyone who directed a specific model, whether it was a person, a group, a university, or a private corporation.

Each of the models was then implemented in two communities and evaluated at the end of one year.

Table 2

Head Start Planned Variation Models[8]

Preacademic Programs

1. *Academically Oriented Preschool Model*, sponsored by Wesley Becker and Siegfried Engelmann, University of Oregon. Emphasis on learning skills in reading, arithmetic, and language through structured drills and reinforcement techniques. Small groups of five to ten children organized into ability levels. Teachers present children with patterned learning materials, elicit constant verbal responses, and provide reinforcement for correct responses.

2. *Behavior Analysis Model*, sponsored by Don Bushell, University of Kansas. Instruction on skills, such as reading and arithmetic, and social skills through reinforcement procedures which feature a token system and individualized programmed instruction. The aim is to have children eventually learn a skill for its intrinsic reward.

Cognitive Discovery Programs

1. *The Florida Parent-Educator Model*, sponsored by Ira Gordon, University of Florida. Involves parents and teachers as educators and thereby insures both home and school instruction. A mother from the local community is a parent-educator who works with parents in their homes and with teachers in school as a teacher's aids. The curriculum is based on Piagetian theories of cognitive development.

2. *Tucson Early Education Model*, designed by Marie Hughes, sponsored by the University of Arizona. Concentrates on language competencies, especially the transition from Spanish to English for Mexican-American children. Allows children freedom to choose activities, fosters cooperation among children, and uses systematic positive reinforcement from teachers in order to promote intellectual, motivational, and social skills.

3. *The Responsive Model*, designed and sponsored by Glen Nimnicht, Far West Laboratory for Educational Research and Development. Curriculum stresses development of a positive self-image and intellectual abilities through self-pacing and self-rewarding materials known as a responsive environment. Materials of instruction consist of problem-solving skills, sensory, discrimination, and language ability. Children freely explore and make discoveries in a structured classroom.

---

[8] Bissell, J.S. Planned variation in Head Start and Follow Through. In J. C. Stanley (Ed.), *Compensatory education for children, ages 2 to 8*. Baltimore: The Johns Hopkins University Press, 1973. Reprinted with modifications by permission.

4. *The Cognitive Model*, by David Weikert, High Scope Educational Research Foundation. Curriculum is based on theories of Piaget. Teachers make decision and plan detailed lessons and activities. Classroom supervisors assist teachers. Home training sessions with mothers are also stressed.

Discovery Programs

1. *A Pragmatic-Oriented Model*, sponsored by the Education Development Center in Newton, Massachusetts. An advisory and consultant system which emphasizes self development and individual needs and styles for teachers and students. Teachers are encouraged to experiment with ways of fostering self respect, respect for others, imagination, curiosity, persistence, openness to change, and ability to challenge ideas.
2. *The Bank Street College Model*, developed and sponsored by the Bank Street College of Education in New York City. Emphasis on each child becoming self-directed in his or her learning. Teachers act consistently as adults children can trust; develop children's sensitivities to sights, sounds, and ideas; and foster children's positive images of themselves as learners. As in traditional middle class nursery schools, children initiate activities, such as cooking, block-building, and sociodramatic activities.

However, the evaluation itself had its limitations: Random assignment of models to communities and children to models and control groups could not be done; consequently, the design was not a classical experimental design, but had all the flaws of a quasi-experimental design (Campbell & Stanley, 1966). Other factors also limited the evaluation: 1) The attrition rate was as high as 50 percent in some sites. 2) The measurement instruments focused primarily on cognitive achievement because instruments that were reliable and valid were available for this outcome; consequently objectives in other domains were assessed by less reliable or valid instruments or were not assessed at all; and even for cognitive achievement, the instruments were weighted for assessing general ability in comparison with general and specific achievenment. 3)Other methodological problems also occurred in the study, such as limitations in the instrument used (floor and ceiling effects); questionable test interpretation procedures, for example, using the child instead of the classroom as the unit of analysis; and inadequate adjustment for initial differences between experimental and control groups. For all these reasons, the results of the evaluation could only be considered for the formulation of hypotheses to be tested subsequently in rigorous small scale experiments rather than as evidence on which definite conclusions could be reached (Bissell, 1973; Kennedy, 1978).[9]

The sample of children consisted of 1,569 in Planned Variation Classes (the experimental group) and 1,078 in "regular" Head Start comparison classes (control group). The children ranged in age from 3 to 6.5 years, but 72 percent were between 4.5 and 5.5 at the beginning of the Head Start Program in the fall. The ethnic sample in percentages were Black, 55; White, 25; American Indian, 2; Puerto Rican, 2; and Mexican-American, 1. Forty-three percent of the parents had attended only grade school and 49.5 percent had attended school only to high school. Although the experimental and control groups were comparable on demographic criteria, marked variation still occurred within specific sites. This variation may not have been well controlled in the analysis of the data.

On the degree of implementation of the models, teachers in the Preacademic model rated highest, followed by those in Cognitive Discovery, and then in the Discovery models. The reason for the variation was not attributable to years of teaching experience, but to the difficulty in learning to implement new types of instructional models. Nevertheless, teachers learned to teach and emphasize their models' areas of concerns. For example, the Preacademic model showed greatest frequency of numbers, letters, and language training activities; Discovery models revealed greatest degree of occurrence of social studies activities and instruction in colors, sizes and shapes; Discovery programs indicated highest amount of role playing activities (doll play and fantasy). However, regular (control) classes also emphasized cognitive training, which would of course, diminish differences between the regular classes and the model classes in this area of instruction.

---

[9] Similar problems were to plague the Follow Through Evaluation as we shall see in the next section.

The results of the comparison between the model and regular classes indicated that the Preacademic and Cognitive Discovery models made the largest gains on academic achievement (a combination of six subtests of the New York University Early Childhood Inventory which assess knowledge related to specific areas of science, mathematics, letters, and numerals) and general cognitive development (the Preschool Inventory and the Stanford Binet Intelligence Scale). However, only 10 Preacademic classes, 8 Cognitive Discovery, and 2 Discovery classes participated in this evaluation. Although children's achievement and cognitive development were positively correlated with the level of academic background and practical experience of their teachers, the same relationship did not occur in the model teachers' backgrounds which suggested that the inservice training conducted by the sponsors made up for these variations among their teachers.

On the Hertzig-Birch measure of styles of coping with cognitive demands, which assesses the way the child responds to tasks on the Stanford-Binet, the results suggested that children in the models programs had learned to engage in task specific activities and to decrease irrelevant verbal and nonverbal responses when asked to do a cognitive task. In short, the children in the models programs seemed to learn to express only correct responses. This type of learning was emphasized in the Preacademic programs, which taught the child to respond only with correct verbal responses. Consequently it is understandable why children in the Preacademic programs made the greatest gains in not responding whenever they apparently could not do the task.

On mother-child interaction, which involved assessing mothers on teaching their children to sort eight blocks and the child's success in doing so and then explaining the basis for the sorting, the greatest gains were made by the Cognitive Discovery and Preacademic classes. Maternal use of praise was highest and children's success greatest in the Preacademic classes. Again, the differential gains among the programs reflected the relative emphases among the models.

When interviewed, parents in regular programs felt that regular Head Start classes provided them primarily with baby sitting and day care for their children. In contrast, parents of Preacademic programs stressed academic and learning performance. In other models, parents stressed interrelationships among children and teachers. Again, the results tended to reflect differential goals of the models.

In general, Bissell (1973) concludes that the findings suggest an *equality of effects pattern*, that is, children tended to have equal gains in cognitive performance and academic achievement in all well-implemented classes. But Bissell also found that a pattern of *specific effects* had emerged: changes in children tended to be consistent with orientation of the program in which the children had participated; these specific effects were greatest on 1) achievement and cognitive measures and 2) cognitive response styles.

Do gains made in preschool programs persist through the primary grades? Citing studies by Beller (1969), Engelmann (1970), Bereiter (1972), and Weikart (1969), Bissell (1973, p. 65) concluded that "poor children who had had preschool experiences perform approximately on a par with their peers who have not had preschool education, and both groups perform below the national norms." Two explanations were given for this phenomenon. They can be called: 1) The *leveling off hypothesis*: Teachers do not build upon gains children made in preschool but instead have them repeat the same curriculum; consequently other children have the opportunity to catch up. 2) The *specificity hypothesis*. Preschool programs only develop functions that are specific to preschool years and these functions do not result in improvement in later years.

However, the failure to maintain gains made in preschool is not exclusive for children from poor and predominantly black families but also from other socioeconomic strata and from Caucasian families as well. For example, Durkin (1974-1975) started a voluntary two-year language arts program for 37 four year olds, 16 girls and 21 boys, in a small midwestern town near Champaign, Illinois, which had some lower and middle class families, but also had a "preponderance of upper-lower class families." All children were white, except for one boy who was black. The mean *IQ* of the group on the Stanford Binet Intelligence Test was 113.6. At the end of two years the average number of words that the 33 students who had continued in the program could correctly identify was 125; they could also name on the average 47 letters (upper and lower case) and they knew sounds for 16 letters. The control group consisted of 26 boys and 23 girls, all Caucasian classmates, who had not attended the preschool but who had attended kindergarten. The control group could identify only 18 words and name 41 letters at the end of the kindergarten year. In grades 1 and 2 the experimental group was significantly above the control group. But in grades 3 and 4, the difference between the experimental and control groups on the Gates-MacGinitie reading achievement test was no longer statistically significant. Durkin (1974-1975, p. 59), implied that the convergence between the two groups had occurred because not all faculty members nor the administrators of the school were "committed to a program in which instruction is matched to children's current achievement. . . ."[10]

At this point, we want to present the results of another attempt to determine whether intensive instruction through a models approach in the primary grades would enable children who had made significant gains in Head Start to maintain their increased rate of development in the primary grades and whether such intervention would enable children from poor families to overcome their educational disadvantagedness. Some studies, for example, Karnes et al. (1969, 1973), had demonstrated that such a continuation would be beneficial. However, in this study, known as Project Follow Through, the models were not only compared with each other, as had been done in the Head Start study, but also with control groups taught by regular public school teachers. A major question then was whether the models approach with experts providing inservice training, curricular materials, and staff help would accelerate educationally disadvantaged students' development.

---

[10] However, precocious readers whose median IQ was 132 were still significantly above a control group on reading achievement (Stanford Intermediate Reading Test) at the end of sixth grade perhaps because the precocious readers' mothers who had played a significant instructional role in their children's achievement were still exerting such an influence (Durkin, 1966).

Project Follow Through

Project Follow Through[11] began as a pilot study in 1967. It was authorized under the Economic Opportunity Act, PL 90-92. The Office of Economic Opportunity then delegated its authority for operating the project to the U.S. Office of Education. Although Follow Through was primarily an educational program, it also provided comprehensive services, including medical, dental, nutritional, social service, guidance, and psychological services to the project's students. Furthermore, parents participated in the program. Consequently, staff members in the project worked not only with schools but also with health, welfare, and social service agencies located in the community. The aim of the project was not just to improve academic achievement but also the child's "life chances." Consequently the project also emphasized other aspects of child development including development of "confidence, initiative, autonomy, task persistence, and health" (Bissell, 1973, p. 88).

The main Follow Through Project began with a meeting in Kansas City in 1968. School district officers who had been invited to the meeting heard descriptions of each model and selected sponsors. A list of the sponsors' models, categorized into three groups according to their major emphases, is in Table 3.

Table 3

Descriptions of Categories and Lists of Models in Each Category[12]

1. *Basic Skill Models*

The models in this category were:

Direct Instruction, University of Oregon

Behavior Analysis, University of Kansas

---

[11] In reporting on Project Follow Through, we have drawn primarily on the four volume report, particularly Volume IV A by Stebbins, St. Pierre, Proper, Anderson, and Cerva (1977).

[12] The description of the models was adapted from Anderson (1977).

Unfortunately the Direct Instruction group adopted a generic name. The group tried to differentiate the project name from the generic name by writing the project name with capital letters in the initial position of each word and the generic name with lower case letters, but it would have been semantically clearer had the group used another name.

Language Development (Bilingual Approach), Southwest Educational Development Laboratory
California Process, California State Department of Education-Division of Compensatory Education

These models focus primarily on instruction in vocabulary, arithmetic, computation, spelling, and language skills. Four subtests of the Metropolitan Achievement Test were used to assess Follow Through effects in these basic skills. The assumption of these models is that the best way to obtain educational effects is through direct instruction, that is, teachers are to teach and have students practice what they want students to learn.

2. *Cognitive-Conceptual Models*
The models in this category were:
Florida Parent Education, University of Florida
Tucson Education Model, Arizona Center for Early Childhood Education
Cognitive Curriculum

These models emphasize "learning how to learn" and problem-solving skills. Tests for assessing these objectives are Raven's Progressive Matrices and Metropolitan Achievement Tests: Mathematical Concepts, Mathematics Problem Solving, and Reading Comprehension. The assumption underlying this model is that if students learn how to learn and how to solve problems, they can apply these processes to acquisition of knowledge in all content areas.

3. *Affective Outcomes Models*
The models in this category were:
Mathemagenic Activities Program, University of Georgia
Responsive Education Model, Far West Laboratory for Educational Research and Development
Interdependent Learning, City University of New York, Institute for Developmental Studies
Bank Street College of Education Approach, Bank Street College
EDC Open Education Follow Through, Education Development Center

These models focus primarily on development of a positive self concept and favorable attitudes towards learning, and secondarily on development of cognitive, conceptual skills. The Coopersmith Self-Esteem Inventory and the Intellectual Achievement Responsibility Scale measure aspects of the affective domain. The assumption underlying these models is that if students accept and feel positive about themselves, they will be able to devote their attention to the external world of learning and achievement.

Table 4 contains a more detailed description of the models.

Table 4
Follow Through Models[13]

Basic Skills Models (Structured Academic Approaches)

These approaches place heavy emphasis on teaching academic skills and concepts within the classroom through programmed instructional techniques. As in the Head Start Preacademic models, each of these approaches uses an analysis of the components which makes up desired behavioral objectives to guide a careful sequencing of learning experiences and a consistent use of external reinforcement. Highly structured educational environments are used by all these sponsors to "engineer" accelerated rates of learning, although they vary among themselves in the specific curriculum content, in the degree of individualized learning, in the respective roles played by teachers, parents, and materials, and in the emphasis placed on the child's initiative and autonomy. The five approaches in this group are:

1. *Behavior Analysis Model*, sponsored by Don Bushell, Jr., Support and Development Center for Follow Through, University of Kansas. In this approach, teachers use a token system of positive reinforcement and individualized programmed materials to teach social skills (such as taking the role of the student) and academic skills in the areas of language, reading, writing, and mathematics; parents are hired to work in the classroom alongside teachers as behavior modifiers and tutors.

2. *Individually prescribed instruction and the primary eduation project (IELP)*, sponsored by Lauren Resnick and Warren Shepler, Learning Research and Development Center, University of Pittsburgh. These approaches provide an individualized, sequenced program of instruction for each child which teaches him academic skills and concepts in the areas of language, perceptual motor mastery, classification, and reasoning. Diagnostic tests determine each child's strengths and weaknesses and are used by the teacher to prescribe instructional materials; positive reinforcement is given continually for success in learning.

3. *The language development/bilingual education approach*, sponsored by Juan Lujan, Southwest Educational Development Laboratory. This approach was originally designed to meet the educational needs of poor Spanish-speaking children (it is currently being adapted for use with French and other non-English-speaking children as well) and teaches mathematics, science, and social studies in the children's native

---

[13] In Bissell, J.S. Planned variation in Head Start and Follow Through. In J.C. Stanley (Ed.), *Compensatory education for children, ages 2 to 8*. Baltimore: The Johns Hopkins University Press, 1973, pp. 90-93. Reprinted by permission.

language, while simultaneously teaching English as a second language; its methods include extensive use of structured drill techniques, reliance on materials relevant to the children's native background and experiences, and development of oral language prior to written language.

4. *The mathemagenic activities program*, sponsored by Charles Smock, School of Education, University of Georgia. Of central importance to this approach is the emphasis on children's learning-by-doing in a sequentially structured environment designed to teach skills and concepts in mathematics, language, science, social studies, art, music, and physical education; children learn through self-initiated, inductive solving of problems which are finely sequenced to assure both advances in understanding and a high level of positive reinforcement.

5. *Direct Instruction*, sponsored by Siegfried Engelmann and Wesley Becker, Department of Social Education, University of Oregon. The primary focus of this program is on promoting skills and concepts essential to reading, arithmetic, and language achievement through structured rapid-fire drills and reinforcement techniques using rewards and praise to encourage desired patterns of behavior; small study groups of five to ten children are organized by teachers according to ability levels in order to facilitate presentation of patterned learning materials and to elicit constant verbal responses from children.

Conceptual Cognitive Models (Discovery Approaches)

The basic goal of the Discovery approach is to promote the development of autonomous, self-confident learning processes in children, rather than simply transmitting specific knowledge and skills. Although, like the Structural Academic approaches, they focus on children's classroom experiences, their emphasis is not on teaching a programmed sequence of materials, but rather on promoting exploration and discovery in an environment which is responsive to the children's own initiative. Heavy emphasis is placed on intrinsic motivation and the gratification children derive from mastery itself. Cognitive growth is seen as only one component of the child's ego development, inseparable from a positive self-concept, curiosity, independence, and the ability to cooperate with others. The three Discovery approaches are:

1. *The Bank Street College model*, sponsored by Elizabeth Gilkeson, Bank Street College of Education. By functioning as consistent adults that children can trust, by being responsive to individual children's needs, and by sensitizing them to sights, sounds, feelings, and ideas, Bank Street teachers help children build positive images of themselves as learners; they introduce themes of study and play relevant to classroom life, encourage children to explore various media, support children's making of choices and carrying out plans and help them use language to formulate ideas and feelings in order to promote self-confidence, environmental mastery, and language expressiveness.

2. *The Education Development Center model*, sponsored by George Hein, Education Development Center. This approach fashions classroom environments responsive to the individual needs and styles of children and teachers in accordance with the "open classroom" concept which has revolutionized British primary schools over the last several years; it is an advisory and consultant system which encourages schools and teachers to experiment with diverse avenues for fostering children's self-respect, respect for others, imagination, curiosity, persistence, openness to change, and ability to challenge ideas.

3. *The responsive environment model*, sponsored by Glen Nimnicht, Far West Laboratory for Educational Research and Development. In this approach, children are free to set their own learning pace and to explore the classroom environment, which is arranged to facilitate interconnected discoveries about the physical environment and the social world. The two primary objectives, helping children develop a positive self-image and promoting their intellectual ability, are achieved through use of self-correcting games and equipment which emphasize problem-solving skills, sensory discrimination, and language ability, and which provide immediate feedback and enjoyment from learning itself.

Affective Outcomes (Cognitive Discovery Approaches)

The Cognitive Discovery approaches are less systematically similar to one another than those in either the Structured Academic or Discovery groups. In general they promote the growth of basic cognitive processes, such as reasoning, classifying, and counting, through highly directed teaching of specific academic skills, through children's autonomous discovery, and through constant engagement of children in verbal activities. Proponents of these approaches share a willingness to be eclectic and to include diverse program elements in their curricula. The four approaches in this group are:

1. *Cognitively oriented curriculum model*, sponsored by David Weikart, High Scope Educational Research Foundation. Derived from the theories of Piaget, this model fosters children's understanding of five intellectual domains (classification, numbers, causality, time, and space) through experimentation, exploration, and constant verbalization on the part of the children, through planning of detailed lessons on the part of the teachers, and through extensive observation and assistance on the part of supervisors. A home-teaching program provides an opportunity for parents to become directly involved in the education of their children.

2. *The Florida parent-educator model*, sponsored by Ira Gordon, University of Florida. In addition to providing ways to improve classroom organization and teaching patterns, this model trains parents to supervise learning tasks in the home in order to increase their children's intel-

> lectual, personal, and social competence. A key element in the program is hiring mothers of Follow Through children as parent-educators who function as teacher aides in the classroom and who work with other mothers in their homes. The curriculum is flexible and varies according to the needs of particular individuals and classes, but there is an orientation toward the theories of Piaget.
>
> 3. *The interdependent learner model*, sponsored by Don Wolfe, New York University. In this model, learning occurs principally in structured small-group instructional "games" where children of different ability levels teach one another and become relatively independent of the teacher. The verbal transactions between children, which are implicit in the process, are a direct stimulus to language development; experiences in phonic blending and decoding skills stimulate reading ability; and language-math-logic games such as Cuisenaire rods and matrix boards promote mathematical understanding.
>
> 4. *The Tucson early education model*, sponsored by Joseph Fillerup, University of Arizona. Major objectives of this model are to promote language competence, intellectual skills necessary for learning (e.g., the ability to attend, recall, and organize), positive attitudes toward school and learning, and skills in particular subject areas, such as reading and mathematics, and in social interaction. Methods emphasize individualized experiences and interest as well as the frequent use of positive reinforcement by teachers.[14]

The result of the meeting was that eventually a total of 170 communities and 75,000 low income children were served on an annual budget of $59,000,000 provided by Congress.

The sponsors and the school districts each had responsibilities for Project Follow Through. Each sponsor was to a) provide a well-defined theoretical model, b) give continuous technical assistance and insure implementation of the model, c) observe progress, d) maintain the model's objectives, and e) participate in evaluation of the model. In turn, the school districts were to a) choose a program sponsor, b) select children for Follow Through who had been in Head Start or similar programs, c) contribute some funds from Title I to the program, d) involve parents and the community in program planning, and e) engage teachers and school staff in the project.

After an initial starting up phase which allowed time for the models to become integrated into school settings, an evaluation plan was initiated in 1971 that used samples of students who had started school in 1970 and 1971 and continued through the third grade.[15] Table 5 lists the grades and years of the Project's evaluation study.

---

[14] Self-sponsored and parent implemented approaches are not included in the analysis in this report; hence they are not described here. For a description of them, see Bissell (1973).

Table 5

Grade Level of Follow Through Children
in Each Cohort and Stream, by Year[16]

	1969-70	1970-71	1971-72	1972-73	1973-74	1974-75	1975-76
Cohort I							
K†	K	1	2	3			
FF††	1	2	3				
Cohort II							
K		K	1	2	3		
FF		1	2	3			
Cohort III							
K			K	1	2	3	
FF			1	2	3		

†Entering kindergarten stream.
††Entering first grade stream.

To make the evaluation fair, tests were selected that reflected the effects of models in each category. The tests that were used are listed in Table 6.

The design of the evaluation was to compare Follow Through Students with a matched control group of non-Follow Through students in kindergarten through grade 3 over the years 1969-1975 for three successive groups of children. But the Follow Through and non-Follow Through groups did not match in economic disadvantage. The median income in 1970 dollars of the Follow Through groups was only $4000 while the control group was $6000.

---

[15] Contracts for the evaluation went to Stanford Research Institute for data collection and to ABT Associates for data analysis. Huron Institute consulted on the evaluation. However, the U.S. Office of Education was responsible for the final design of the evaluation.

[16] Stebbins, L.B. et al. An evaluation of Follow Through, Vol. IV A. In R.B. Anderson (Project Director), *Education as experimentation: A planned variation model* (Final Report). U.S. Office of Education, 1977.

Table 6
Instruments used in Assessing Objectives in Grades K-3

Objectives	Tests	Entering	Leaving			
		K or 1	K*	1*	2*	3
Basic Skills						
	Wide Range Achievement Test	♦ or ♦				
	Peabody Picture Vocabulary	♦ or ♦				
	Caldwell Preschool Inventory	♦ or ♦				
	Metropolitan Achievement Test (MAT)		♦			
	Listening for Sounds					
	MAT-Primary I or II or Elementary**					
	Word Knowledge			♦	♦	♦
	Spelling			♦	♦	♦
	Math Computation			♦	♦	♦
	Language Arts (A and B)					♦
Cognitive Conceptual Skills						
	MAT-Primer**, Primary I, II, Elementary					
	Reading		♦	♦	♦	♦
	Math Concepts		♦	♦	♦	♦
	Math Problem Solving				♦	♦
	Raven's Coloured Progressive Matrices					♦
Affective Skills						
	Intellectual Achievement Responsibility Scale					
	Locus of Control (positive)		♦			♦
	Locus of Control (negative)		♦			♦
	Absence from School		♦			
	Coopersmith Self Esteem Inventory			♦		♦

*Measures used only for the third successive group, Cohort III who started in kindergarten (IIIK).

**MAT-Primer was given at end of kindergarten, Primary I in grade 1, Primary II in grade 2, and Elementary in grade 3.

The experiment also had other limitations:

1. *Specificity of treatment.* a) Changes continued to occur in models' objectives and goals and b) there was no adequate assessment of degree of implementation of each model.

2. *Measurement battery.* Basic skills domain was represented better in the measurement battery than were the conceptual and affective categories of models; and

3. *Weakness in design.* a) No random assignment to treatment of subjects or sites and b) the Follow Through and non-Fol-

low Through groups were not matched on the economic criterion. However, the failure in matching favored the control group and therefore biased the results against the experimental group (Follow Through).

The ethnic distribution of the students was markedly skewed. Blacks and Puerto Ricans were predominantly in large cities, Mexican-American pupils were concentrated in the West, and white students were most frequently in the North Central areas.

Three types of comparisons were made:

1. *Local.* Follow Through students were compared with non-Follow Through students at local sites.

2. *Pooled.* Follow Through students were compared with pooled or total non-Follow Through students.

3. *Norm referenced.* Follow Through students were compared with standardized test norms.

These comparisons were designed to answer these questions:

1. Does Follow Through have a greater educational effect on disadvantaged children than does the regular school program?

2. Do the various models have differential effects on the education of disadvantaged children?

3. Which, if any of the models, show consistent effects over time?

Underlying the questions were these hypotheses:

1. *Cultural deprivation.* In contrast to middle class children, educationally disadvantaged children do not receive as much formal or informal preschool education. Consequently the cognitive and affective aspects of development are adversely affected. Since these aspects of development underlie educational achievement, their school performance is also inadequate. Some Follow Through models focused on compensating for this early deprivation by attempting to accelerate the rate of cognitive and affective development of Follow Through children.

2. *Failure of schooling.* The relatively low achievement of educationally disadvantaged children is attributable to teachers

who have low expectations for poor children and consequently adapt the curriculum downward, exhibit negative attitudes toward these children, and use ineffective instructional procedures and techniques in teaching them. The antidote is better teachers and teaching with higher academic expectations, more systematic instruction with a challenging and stimulating curriculum, and a more favorable attitude towards disadvantaged children in the classroom.

The results of the study are voluminous. Fortunately they have been summarized in three charts. Figure 2 depicts the average effects on Basic Skills, Figure 3 on Cognitive-Conceptual Skills, and Figure 4 on Affective Outcomes. These graphs were constructed by determining whether the Follow Through group at a particular site had a significant positive or negative effect on one of the instruments as compared with its matched non-Follow Through group. A positive or negative effect was significant if the difference between the two groups on the measure was at least one quarter of a standard deviation difference. This difference was about equal to two months in grade placement on a standardized test. Then an average effect was computed for each site according to the following formula:

$$\textbf{Average effect}=\frac{\textbf{Number of positive effects}-\textbf{Number of negative effects}}{\textbf{Total number of effects}}$$

The results were then plotted to make the bar graph. For example Figure 2 shows that the Direct Instruction model for Follow Through children on the Basic Skills category had more sites in which there were positive than negative effects. The mean of the magnitude of positive and negative effects, as indicated by the dot on the graph for the Direct Instruction model, is about 0.4. In other words, when compared with non-Follow Through classes at its various sites, the Direct Instruction model came out more often on the positive than on the negative side of the comparison. However, note that there was considerable variation for each model.

Figure 2
Ranges of Site Average Effects on Basic Skills for Thirteen Follow Through Models[17]

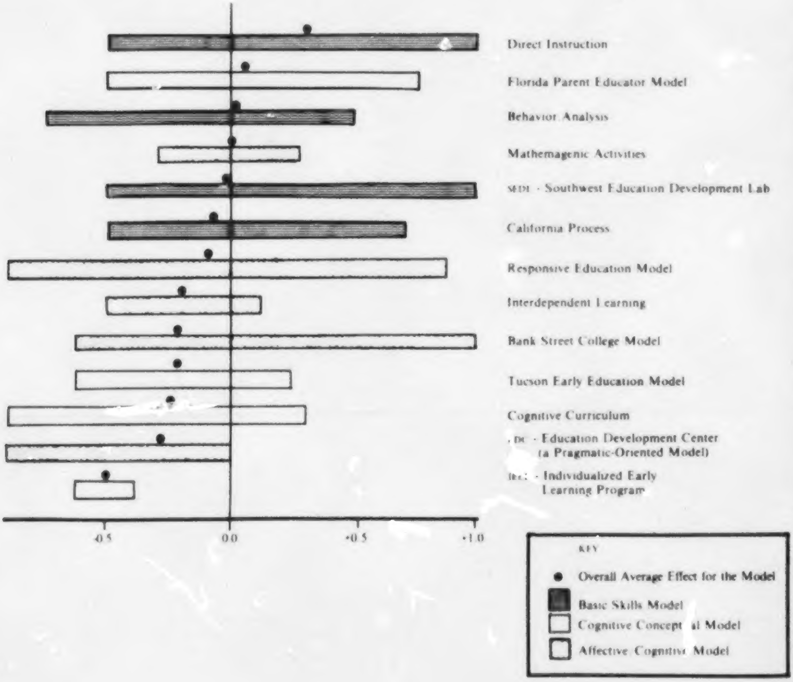
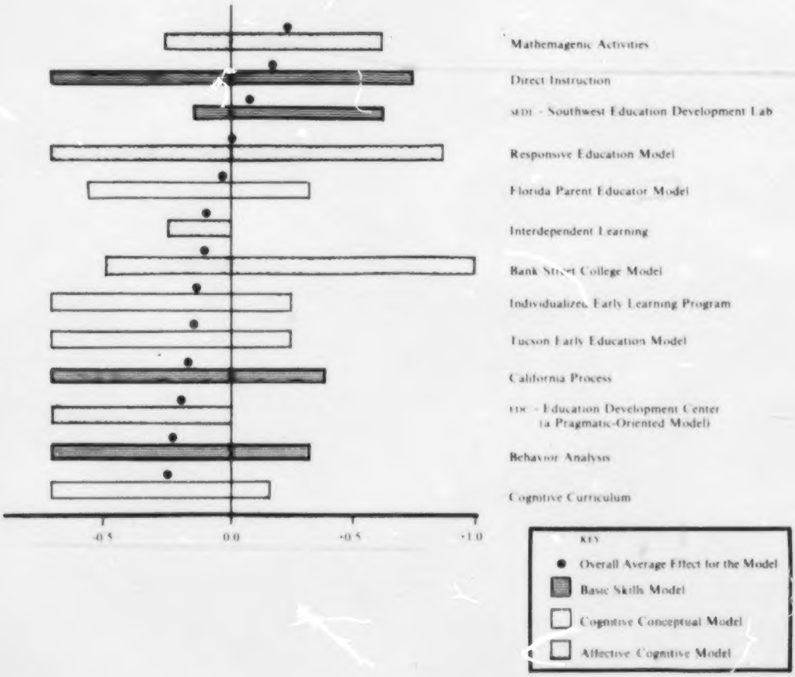


Figure 2 also shows that the Direct Instruction model made on the average greater gains in Basic Skills when Follow Through students were compared with non-Follow Through at its sites than did any of the other models, as shown by a comparison of the dots in each graph. The graphs in Figures 3 and 4 are to be read in the same way. Note that on Figure 3, Cognitive Conceptual Skills, Direct Instruction came out second; and on Figure 4, Affective Outcomes, Direct Instruction was again first in comparison with the other models.

---

[17] Stebbins, L.B. et al. An evaluation of Follow Through, Vol. IV A. In R.B. Anderson (Project Director), *Education as experimentation: A planned variation model* (Final Report). U.S. Office of Education, 1977.

Figure 3
Ranges of Site Average Effects on Cognitive Conceptual Skills for Thirteen Follow Through Models[18]



A succinct summary of the results which reflect "clear and general patterns" has been prepared by Anderson (1977, pp. *XXIV-XXIX*). Since we find their summary does agree with the data we shall follow it in presenting the results:

Variability among Models (Figures 2, 3, 4)

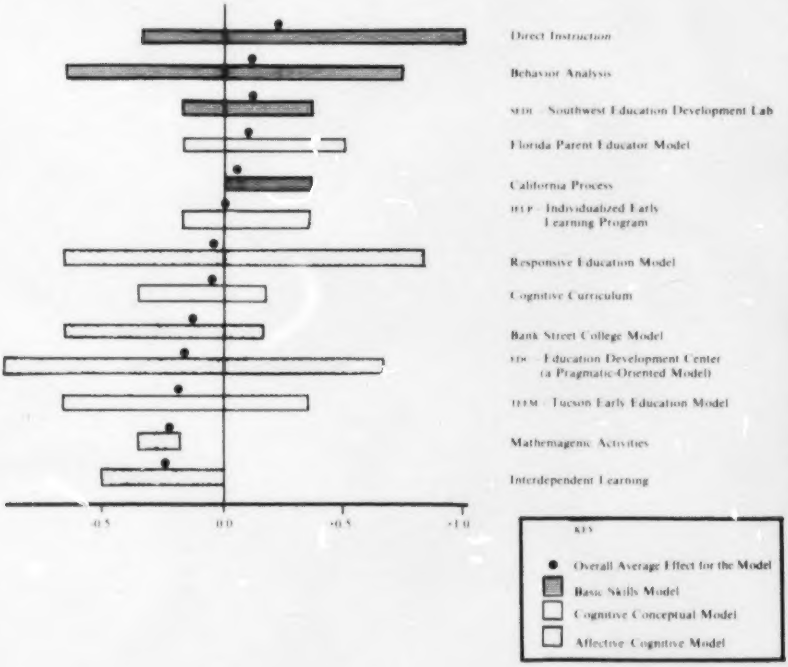
Models varied in their effectiveness from site to site.

In some sites Follow-Through models had more positive than negative effects while in some other sites non-Follow

---

[18] In Linda B. Stebbins et al. An Evaluation of Follow Through. Volume IV A. In Richard B. Anderson (Project Director), *Education as Experimentation: A Planned Variation Model*. Final Report, U.S. Office of Education, 1977.

Figure 4
Ranges of Site Average Effects on Affective Outcomes for Thirteen Follow Through Models[19]



Through groups had superior performances. No model was successful in every site nor did any model fail in every site on Basic Skills and Cognitive-Conceptual Skills. However , seven models which stressed Affective Outcomes performed on the average more poorly in each type of assessment than did their comparison non-Follow Through groups, as shown by the number of dots on the negative side of the bar graphs in Figures 2, 3, and 4.

Basic Skills Domain (Figure 2)

1. *Models that stress basic skills are more effective in developing these skills than other models which emphasize cogni-*

---

[19] In Linda B. Stebbins et al. An Evaluation of Follow Through, Volume IV A. In Richard B. Anderson (Project Director), *Education as Experimentation: A Planned Variation Model.* Final Report, U.S. Office of Education, 1977.

tive-conceptual abilities or affective outcomes.. In short, to improve basic skills, teach them directly rather than rely upon transfer from improved cognitive-conceptual abilities or affective outcomes. In relation to the two hypotheses on which the study was based, the failure of schooling hypothesis gains support from the results attained by the Basic Skills models . However, it should be noted that two of the four basic skills models were generally less effective in developing basic skills than were their control groups.

2. *Regular school instruction was more effective than most models in developing basic skills.* By grade 3, basic skill development was greater for non-Follow Through sites than Follow Through sites for all models except Direct Instruction, Parent Education, Behavior Analysis, and Mathemagenic Activities, perhaps because the other models diverted children's time away from the basics. Evidence is mounting that "time on task" is a major determinant of achievement related to educational tasks (Rosenshine & Berliner, 1978).

If "academic engaged time" (Rosenshine & Berliner, 1978) is a significant factor in teaching basic skills, then in the non-Follow Through sites basic skills apparently were stressed more in non-Follow Through groups than they were in groups taught by the models, particularly the Cognitive-Conceptual and Affective models. Moreover, one-third of the sites even exceeded the performance levels attained by the Direct Instruction model. Apparently, the specialized materials and procedures claimed to be among the effective ingredients in the Direct Instruction model were not as effective as the materials and procedures used in one-third of the regular or non-Follow Through classes with which the Direct Instruction classes were compared.

Cognitive-Conceptual Domain (Figure 3)

No one category of models was more successful than another category in raising performance on cognitive-conceptual skills. Indeed those models which purported to stress cognitive-conceptual skills tended to perform not as well as models in other categories did in raising Follow Through children's scores when the scores achieved by their groups are compared with non-Follow Through children. In fact, it will be noted that for nine of the thirteen models, control group (regular classroom) pupils made greater gains than did the experimental group children.

Affective Domain (Figure 4)

Basic Skills Models produced performance in the affective domain that was as high or higher than performance attained by models that stress the affective domain. Apparently the Basic Skills Models did not achieve their gains at the expense of an adverse effect upon children's self concepts. An hypothesis which appears to be plausible is that as a consequence of improvement in basic skills in the primary grades, a child's self concept is enhanced. Again it should be noted that for seven of the thirteen models, control group children performed higher on the affective outcomes measures than did the experimental group children.

Conclusions

We have briefly reviewed the Coleman Report which provided a ground plan for the development of compensatory education programs following the 1954 Supreme Court decision on equality of educational opportunity. We have also reviewed the two major federally-funded approaches to compensatory education: Project Head Start, and Project Follow Through. We have looked at these recommendations and projects in light of practical school programs, such as the voluntary desegregation program in the Riverside Unified School District, and the Durkin precocious reader study, with an eye to arriving at conclusions which might guide the development of instructional programs in public schools. Recognizing the limitations in the data, we believe the following conclusions are warranted.

The preeminent conclusion which may be drawn from the Project Follow Through evaluation is that the ease of improving elementary education in the United States has been grossly underestimated by the United States Office of Education, the Congress, and most of the thirteen groups of curriculum developers and researchers responsible for the experimental programs tested in the project. Control group classrooms in the preponderance of comparisons scored higher on the dependent variables than did the experimental classrooms. They scored higher on basic skills than nine of the thirteen model groups; they scored higher on Cognitive-Conceptual skills than nine of the thirteen model groups; they scored higher on affective outcomes than seven of the thirteen model groups. Regular classroom teachers teaching the regular curriculum of their school, without the benefit of the additional millions of dollars of federal funds, outperformed the experimental groups in all outcome categories.

Perhaps these findings support the belief that common school practice in the United States today represents the collective wisdom from more than one hundred years of pedagogical experience, and improvement of educational outcomes will depend more upon outcome from specific instruction than upon general approaches to general objectives.

A second conclusion from this review is that if you want pupils to learn a particular skill or knowledge it is more efficient to teach it directly than to expect it to transfer from other learnings. As a group, the Basic Skills Model programs (or Direct Instruction model) were superior to either of the other two groups on the basic skills measures. The program in this general model taught the specific skills required for word analysis and word recognition and the pupils learned these skills better than they were learned by pupils in the other model programs.

Third, though the Affective/Cognitive Model programs were directed more toward achieving affective outcomes, the Basic Skills Model programs enjoyed their greatest superiority on these dependent variables. Only in this area were the Basic Skills approaches consistently superior to the control group approaches as well as being superior to all but one of the other model programs. It seems reasonable that improved performance on affective outcome measures will be directly related to increased achievement in cognitive areas when programs focus upon the development of cognitive skills and knowledge.

A substantial part of the life of the child is spent in school. In the school setting, parents prize academic achievement, teachers demand academic achievement, and pupils realize that somehow their status is dependent upon academic achievement. Basic skills instruction secures the greatest academic growth, and fosters high affective outcomes.

References

ANDERSON, J.B. (Project Director). *Education as experimentation: A planned variation model* (Final Report). Washington: U.S. Office of Education, Department of Health, Education, and Welfare, 1977.

ANDERSON, R.B., ST. PIERRE, R.G., PROPER, E.C., & STEBBINS, L.B. Pardon us but what was the question again? A response to the critique of Follow Through evaluation. *Harvard Educational Review*, 1978, *48*(2), 161-170.

BELLER, E.K. The evaluation of effects of early educational intervention on intellectual and social development of lower class, disadvantaged children. In E. Grotberg (Ed.), *Critical issues in research related to disadvantaged children*. Princeton, N.J.: Educational Testing Service, 1969.

BEREITER, C.E. An academic preschool for disadvantaged children: Conclusions from evaluation studies. In J.C. Stanley (Ed.), *Preschool program for the disadvantaged*. Baltimore: The John Hopkins University Press, 1972.

BEREITER, C., & ENGELMANN, S. *Teaching disadvantaged children in the preschool*. Englewood Cliffs, N.J.: Prentice-Hall, 1966.

BISSELL, J.S. Planned variation in Head Start and Follow Through. In J.C. Stanley (Ed.), *Compensatory education for children, ages 2 to 8*. Baltimore: The Johns Hopkins University Press, 1973.

CAMPBELL, D.T., & STANLEY, J.C. *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally, 1966.

COLEMAN, J. The concept of equality of educational opportunity. *Harvard Educational Review*, 1968, *38*(1), 7-22.

COLEMAN, J.S. Recent trends in school integration. In G.V. Glass (Ed.), *Evaluation studies: Review annual* (Vol. 1). Beverly Hills, Calif.: Sage Publications, 1976.

COLEMAN, J.S., CAMPBELL, E., HOBSON, C.J., MCPORTLAND, J., MOOD, A.M., WEINFELD, F.D., & YORK, R.L. *Equality of education opportunity*. Washington: Office of Education, U.S. Department of Health, Education and Welfare, 1966.

DURKIN, D. The achievement of pre-school readers: Two longitudinal studies. *Reading Research Quarterly*, Summer 1966, *1*(4), 5-36.

DURKIN, D. A six year study of children who learned to read in school at the age of four. *Reading Research Quarterly*, 1974-975, *10*(1), 9-61.

DYER, H.S. The measurement of educational opportunity. In F. Mosteller & D.P. Moynihan (Eds.), *On equality of educational opportunity*. New York: Vintage/Random House, 1972.

ENGELMANN, S. The effectiveness of direct instruction in IQ performance and achievement in reading and arithmetic. In J. Hellmuth

(Ed.), *Disadvantaged child* (Vol. 3). New York: Brunner-Mazel, 1970.

FOWLER, W.F. *Concept learning in early childhood.* Unpublished paper presented at the annual meeting of the American Educational Research Association, February 10-12, 1965.

GUTHRIE, J.T., SAMUELS, S.J., MARTUZA, V., SEIFERT, M., TYLER, S.J., & EDWALL, G. *A study of the locus and nature of reading problems in the elementary school.* Final Report, National Institute of Education, 1976. (ERIC Document Reproduction Service No. ED 127 568)

HENDRICK, I. *Public policy toward the education of non-white minority group children in California, 1849-1970.* National Institute of Education, Project No. NE-G-00-3-0082, Final Report, 1975.

HOUSE, E.R., GLASS, G.V., MCLEAN, L.D., & WALKER, D.F. No simple answer: Critique of the Follow Through evaluation. *Harvard Educational Review*, 1978, *48*(2), 128-160.

JACKSON, G. Reanalysis of Coleman's "Recent trends in school integration." In G.V. Glass (Ed.), *Evaluation studies: Review annual* (Vol. 1). Beverly Hills, Calif.: Sage Publications, 1976.

KARNES, M.B. Evaluation and implications of research with young handicapped and low income children. In J.C. Stanley (Ed.), *Compensatory education for children, ages 2 to 8.* Baltimore: The Johns Hopkins University Press, 1973.

KARNES, M.B., HODGINS, A.S., TESKA, J.A., & KIRK, S.S *Research and development program on preschool advantaged children. Investigation of classroom and at-home interventions.* Washington, D.C.: Office of Education, U.S. Department of Health, Education, and Welfare, 1969.

KATZ, I. Academic motivation. *Harvard Educational Review*, 1968, *38*, 57-65.

KENNEDY, M.M Findings from the Follow Through planned variation study. *Educational Researcher*, 1978, *7*(6), 3-11.

PENA VS. SUPERIOR COURT OF KERN COUNTY (1975) 123 Crt. Rptr. 500, 50 C.A. 3d 694. Summarized in West's Ann. Calif. Educ. Code #35351, 1978.

PINES, M. A pressure cooker for four-year-old minds. *Harper's Magazine*, 1967, *230*(1400), 55-61.

PURL, M.C., & KLEINKE, C. *Comparative data for the Lorge-Thorndike Intelligence and Stanford Reading Achievement Test.* Riverside, Calif.: Department of Research and Evaluation, Riverside Unified School District, September 1968. (multilithed)

ROSENSHINE, B., & BERLINER, D.C. Academic engaged time. *British Journal of Teacher Education*, 1978, *4*, 3-16.

SINGER, H. *Effect of integration on achievement of Anglos, Blacks, and Mexican-Americans.* Paper read at the Annual Convention of the American Educational Research Association, Minneapolis, March 3-6, 1970.

SINGER, H., GERARD, H.B., & REDFEARN, D. Achievement. In H.B. Gerard & N. Miller (Eds.), *School desegregation: A long-term study.* New York: Plenum Press, 1975.

SINGER, H., & HENDRICK, I.G. Total school integration: An experiment in social reconstruction. *Phi Delta Kappan,* 1967, *49*(3), 143-147.

STEBBINS, L.B., ST. PIERRE, R.G., PROPER, E.C., ANDERSON, R.B., & CERVA, T.R. Volume IV A: An evaluation of Follow Through. In R.B. Anderson (Project Director), *Education as experimentation: A planned variation model.* Final Report, U.S. Office of Education, Department of Health, Education, and Welfare, 1977.

WEIKART, D.P. *A comparative study of three preschool curricula.* Paper presented at the annual meeting of the Society for Research on Child Development, Santa Monica, California, 1969.

WILSON, A.B. Social stratification and academic achievement. In A.H. Passow (Ed.), *Education in depressed areas.* New York: Teachers College, Columbia University, 1963.

WISLER, C.E., BURNS, G.P., JR., & IWAMOTO, D. Follow Through redux: A response to the critique by House, Glass, McLean, and Walker. *Harvard Educational Review,* 1978, *48*(2), 171-185.

How to Recognize a Reading Comprehension Program: An Afterword

John T. Guthrie
International Reading Association

We have enormous powers of recognition. For most people, it is nothing to recognize a hospital, a farm, a game of chess, an oak tree, or a reading program. We can distinguish a hospital from a bank, a farm from a hill, chess from football, a tree from a bush, and a reading program from a concert. We possess little awareness, and little need for awareness, about what we retain in our mental structures that allows us to make these perceptions rapidly, automatically, and thousands of times a day.

When it comes to reading programs, however, professionals in the field need to be able to reliably go beyond the correct identification. A bird watcher may correctly identify a thrush, a robin, or a stilt without being able to name their unique features. Yet, the difference between the ornithologist and the amateur is that the ornithologist has a store of knowledge that can be communicated to other people which will enable them to discriminate these species from each other and to describe the biological difference between birds and non-birds.

The challenge for students of reading is not to recognize a reading comprehension program when it is presented, nor to have the concept generally developed in the head, but, rather, to externalize the knowledge in a way that will allow another person to make the recognition and to distinguish a good example from a poor one.

Although the exercise of recognition is adequate for communicating about reading programs among insiders, it is not adequate for other purposes. For example, perpetuating programs that currently exist requires intelligence as well as inertia. Education of reading professionals involves the conveyance of knowledge, as well as the trial and error of practice teaching. The improvement of reading education programs demands criteria of excellence, rather than conformity to current fashion.

To construct a system that will recognize a reading program, we may ask about the generally accepted meaning of "a reading program." Typically, definitions have centered on the materials children are expected to read. In 1967, Chall reported her "critical analysis of existing research comparing different approaches to beginning reading" (p. 5). She agreed with Barton and Wilder (1964) that, "Reading instruction in almost all schools starts from a similar basis: basal readers from a graded series are used by 98 percent of first grade teachers and 92 to 94 percent of second and third grade teachers on all or most days of the year" (p. 188). Due to the dominating influence of the basal reading series, Chall went to some length to describe one of them and gave the following definition of series in general.

> A basal reading series...attempts to give teachers and pupils a "total reading program" embodying a system for teaching reading (in the teachers' manuals), a collection of stories and selections for pupils to read (the readers), and exercises for additional practice (the workbook). The series generally starts with a prereading program (one, two, or three "reading readiness" books for kindergarten and/or grade one). Then come the "graded" readers: three or more small paper-covered books (the preprimers), which are followed by the first hard-covered book (the primer), and then a first reader (the 1-2 book). Typically, these five (sometimes six) books are used by children in the first grade, although some advance through them faster and some slower. The typical basal reading series has a book for each half of the second grade—the 2-1 and 2-2 books. Similar breaks are made for the third through the sixth grades. Some companies have readers for the seventh and eighth grades of elementary school or the ninth grade of junior high school. With each reader goes a workbook—a "consumable," paper-covered book containing a variety of exercises. Each reader also has a teacher's guidebook (a manual on how to teach from the readers and workbooks). Schools can also purchase accompanying charts, tests,

> and other aids with each series. In addition, some publishers offer supplementary readers and subject matter textbooks coordinated with the basal reading program (p. 188).

This emphasis on materials was not altered when Chall interviewed the authors of major basal systems prominent at that time. Most authors recommended that the program be supplemented by related reading from the library, children's literature anthologies, and textbooks. Otherwise, elaborations on what a reading program might be were not given.

Another illustration of the prevailing orientation to materials is given by Bond and Dykstra (1967) in their description of the reading programs that formed the basis of the first grade reading studies. They noted that

> A fourth treatment group was labeled Language Experience. A basic element of this instructional method is that the child's own writing serves as a medium of instruction. The child's first stories are dictated to the teacher who acts as the recorder. As soon as he is able, the pupil writes his own stories and shares them with the teacher. During the individual conferences between pupil and teacher, he is helped to recognize the commonality between the words he writes and speaks and he develops the skills necessary for reading. This approach, then, ordinarily utilizes far fewer highly structured instructional materials than do most instructional programs. In addition, vocabulary control is viewed as being in the language itself and in the language background of each child. The pupil learns to read the words necessary for him to use in writing. One of the major instructional tasks in this method is to engender a stimulating language environment (p. 46).

In this description, the content to be read by the child (the child's dictated materials) is emphasized. The content is the dominant feature of this reading approach according to these authors.

The astonishing fact about these two definitions is that they almost totally ignore the student as a learner. It is as though a reading program could somehow exist independent of the children expected to learn from it. If teachers in an elementary school in the United States are asked what reading program they use, their replies are likely to be the names of such and such materials—usually basal reading series. It goes without saying that these materials are not simply warehoused or used as doorstops; they are processed by children. Although children learn from the materials, this learning is so inherent that it often goes without notice or description. Yet a reading program is more than a set of materials or books or stories or exercises. A reading program is a collection of contents to be read, studied, pondered, answered, and learned.

More recently, the concept of a reading program has included statements of objectives about what children are expected to learn. For example, a contemporary reading program author states in the teacher's edition that seven skill areas permeate the program. These include: 1) decoding, which "may be viewed as the conversion of written symbols into the sounds they represent"; 2) comprehension, "the ability to grasp the meaning or meanings of communication"; 3) language, which refers "to the child's production of language and his/her understanding of language itself"; 4) research and study skills, which includes "finding and using various kinds of information"; 5) literature and the writer's craft, which is "concerned with understanding different kinds of literature, distinguishing between fact and fiction, appreciating an author's style, etc."; 6) attitude toward self in reading, which is the way children react to reading on a personal basis; and 7) creativity, which is the encouragement of new, different, and unusual responses. Although these objectives may or may not be accomplished and the new materials may or may not be suitable for them, they stand as goals which are related to student learning and comprise the intention of the program (Ruddell & Crews, 1978).

A recent, extensive analysis of the Houghton Mifflin and Ginn 720 Series (Beck, McKeouen, McCaslin, & Burkes, 1979), states that these series provide "total instructional systems," including the pedagogy for teaching selections for children to read and practice exercises. Beck, et al. claim the programs "exhibit similarly broad orientations to reading, which include decoding, literal and inferential comprehension, and enrichment skills" (p. 5). They point out that most units (in other words, the nature of the program) are conditioned by the nature of the objectives for learning that accompany the materials and are embodied in them.

The recent addition of objectives to the concept of a reading program is still not sufficient to represent what people really mean by a reading program. If a teacher stood in a classroom equipped with a basal reading series, and there were no children in the classroom for one year, one would hardly say that a reading program had occurred. Likewise, if the children went to the classroom daily but never looked at the available basal materials nor any other reading materials, one would hardly say that a reading program had been experienced by the students. If the children looked at the materials but paid no attention in the extreme or found them to be written in a foreign language, one would not say the reading program had taken place. In the end, some learning must occur for at least some children to allow the inference that a reading program is being conducted in a classroom. Otherwise, no program exists.

We have seen that processes of learning in children are intrinsic to the body of a reading program. These processes may include attention, visual and auditory perception, language comprehension, linguistic differentiation, understanding word meanings, and written communications. The point is that to see or describe or account for a reading program is to contend not only with the materials that are intended to produce learning but to capture the learning or information processing or reading comprehension that is intended to occur as a response to the materials provided. Without describing the kinds and amounts of comprehension that occur, we can't affirm that a program exists at all, much less whether it has been implemented as it was intended. To recognize a reading program, one must be able to recognize reading comprehension and its improvements in the learner.

The present book contains several chapters of research on reading in the 1970s that are directed to describing the processes of reading comprehension. For example, Rumelhart illustrates that a schema is a representation of knowledge. In his network are included units of information with relationships among them in a certain pattern. From a schema, one constructs an interpretation of incoming information from written materials.

These schemata, then, determine what can be understood during the process of reading. In addition, schemata facilitate the recall of stories and text materials. If a person has a well-developed network of relationships in memory for a particular content area, then what one reads that is related to the schema will be clearly retrievable for relatively long periods of time.

To recognize that reading comprehension is occurring in a classroom for a child, we must know the type and refinement of schemata possessed by the child, the type of text being read by the child, and the likely interaction between the two. If the child answers correctly a question over a paragraph, it is possible that comprehension did not occur because the child already possessed the information in the form of background knowledge. It is possible that comprehension did not occur because the child answered the question "rotely" from the text. Furthermore, if the child fails to answer correctly, it is possible that it is not a problem of comprehension but rather one of insufficient background information, inappropriate vocabulary in the text, inadequate attention to the text, or unwillingness to provide an answer that is available in the child's response system. Therefore, to recognize comprehension we must perceive a correct answer to a question over material that was read by the child that could not have been answered from background knowledge or verbatim recall of the material in the text. To detect improvements in these abilities is to perceive correct answers to questions over text in which the text has increasingly complex language structures and the knowledge base of the child is larger in quantity and more complexly related to the text.

In the present volume, Pearson and Camperell have illustrated that children possess not only schemata for content of written materials, but, also, schemata for structure. For instance, a child's story is likely to have a set of syntactic relationships at the level of the story consistent with many other narratives. Pearson suggests that a schema for stories can be taught to elementary school children. If this is true, then one may recognize the acquisition of comprehension in a classroom if a child's schema for narrative is improving. A "better" schema is one that contains more adultlike constituents and their relationships. An increasingly complex schema allows children to understand and remember increasingly complex narratives and to identify finer distinctions within them, with respect to the plot, its episodes, characters, resolution of conflict, and attitude of the writer.

One process necessary to comprehension is that of making inferences. As Trabasso (in this book) has shown, inferencing allows the reader to connect the text in such a manner as to render a plausible interpretation for its meaning. Without a host of inferences, most text is fragmented to the point of incoherence. There are several types of inference—causal, informational, temporal, evaluative, motivational, and so on. Trabasso emphasizes that success in answering causal inference questions is crucial to story comprehension. In his chapter there is evidence that children who are better comprehenders are more likely to answer causal questions correctly than are children who are inferior in this process. This implies that one aspect of improving comprehension is the improvement of inferences in terms of frequency, complexity, and appropriateness to the material for which suitable schemata are available.

Although word meanings are highly correlated to reading comprehension and a substantial number of investigations have been made into the relationship (over 3,000 since 1972 by one count), the reason for the association is not clear. As Anderson and Freebody suggest here, it may be that these variables are correlated because word meanings are part of the task of reading comprehension. Another possibility is that vocabulary is a measure of aptitude for which comprehension is another indicator. It is also possible that vocabulary assesses the knowledge base that is indispensable for comprehending a paragraph written on that topic.

Equally mysterious, according to Anderson, is just what it means to "know the meaning of a word." To simply know its denotative meaning will be inadequate for many purposes, and yet it is not possible for young children to know its subtle connotations that may come out in a variety of contexts. The size of a student's oral vocabulary involves both depth and breadth of word knowledge in a complex relationship. Despite the philosophical complexity and empirical uncertainty in this area, it seems justified to recognize improvement in reading comprehension when children's word meanings increase in number and complexity, and words that previously rendered a sentence or phrase incomprehensible do not present the same obstacle at a later point in time.

There is one definition of a reading program that focuses on the processes of comprehension and learning. Stauffer (1969) states that

> The objective of group (directed reading teaching activity) training is to develop skill in critical reading. To be a critical reader requires a command of three intellectually refined skills. The first skill is the art of inquiry, or of asking relevant questions. Allied with this is the ability to conjecture, estimate, hypothesize. A scholar does not raise questions in an intellectual vacuum; rather by virtue of the knowledge and experience available to him at the time, he conjectures about answers. If he accepts the questions that someone else raises, it is even greater significance that he speculate about answers. The second skills required is that of processing information. The scholar does this in a prescribed, not random, order, as dictated by his reading goals. In many ways, the human mind acts like a computer if properly programmed (inquiry-regulated, goal-set) data encountered are processed appropriately (as when evidence is weighed selectively). The third skill is that of validating answers. This the scholar does either by testing his judgment against that of a group or by actually trying the answers to see if they are correct. Thus, in brief, the critical reader first feeds forward, or inquires, then he processes selectively, and finally he gets feedback by testing answers (p. 35).

In this view, it is not the materials nor their objectives that are the defining quality but, rather, the processes performed by the learner. The teacher's responsibility is to provide written stimulation and sufficient interaction with children to enable these cognitive accomplishments to take place.

A natural question at this point is, "What characteristics of the reading materials will set into motion comprehension processes and generate the conditions for their improvement?" An attempt to answer this was made by Jenkins and Pany (in this volume). A survey of the experimental literature in instructional psychology revealed that the following characteristics have been illustrated to improve comprehension for written materials: appropriate headings placed in expository text, pictures that give supplemental information to adjunct material, speed and fluency of word recognition, instruction in word meanings that are explicitly required by sentences, questions in the text that are directly related to information to be learned, and mental elaborations on what is comprehended from reading.

The strategy employed by Jenkins and Pany in their review was to locate variables that have been found to correlate with comprehension and examine whether improving performance on these variables through intervention increases comprehension ability measured in another, appropriate form. Such synthesizing strategy is painstaking and requires sophistication in experimental methodology. However, the findings are well rooted in a conceptual base of instructional psychology and may serve as benchmarks for promising educational materials. A reading program that contains them is far more likely to evoke constructive comprehension processes than one in which they are missing or sparsely scattered.

In the reading program definitions given by Chall, Bond and Dykstra, Beck, and the authors of major published curriculum series, little is said about quantity of instruction. Yet, suppose a teacher had a desirable set of materials well suited to the children but only engaged the children in a lesson over reading for one day of the school year. In practical terms, the school that teaches math 179 days and reading for 1 day doesn't have a reading program. Yet, what about 2 days of reading instruction? Is that enough? How about 2 months? How many hours of reading activity is needed to conclude that a reading program has taken place within a school? This overwhelmingly obvious point has been neglected in the '50s, '60s, and '70s in the quest for an optimal method. What is being asked now is, "What is the optimal amount?"

Berliner introduces the concept of academic engaged learning time to describe the quantity of instruction. It is based on the simple fact that the best predictor of reading achievement is the amount of time spent reading during the course of the school year. This variable has been identified by Harris and Serwer (1966) and highlighted by Rosenshine (1978), but it has been experimentally investigated by Berliner. He reports here that the range in amount of time spent engaged with materials of an appropriate level is broader than one might suppose: from 24 to 59 minutes a day, which is from 60 to 148 hours per year spent in reading instruction by children. Berliner also illustrates that if children read materials which are appropriate in difficulty for as little as 4 minutes a day for a period of 5 weeks in the school year, their reading achievement will not change, which means their percentile rank will decrease by 10. In his study, if children engaged in reading 23 minutes a day for 5 weeks, their percentile rank stayed the same. And if 52 minutes a day were spent in appropriate reading activities, an increase of 15 percentile points was observed in a 5-week period. What these data point to is that to recognize a reading program, one must measure its quantity.

Berliner's analysis underscores the fact that the amount of a reading program is indexed by the amount of time children spend learning, rather than the amount of time allocated by the teacher or the amount of time the materials may be available in the library for use or the amount of time teachers may spend planning. It is noteworthy that for many reading educators, Sustained Silent Reading (SSR) has been hailed as a breakthrough. Across the country, teachers have taken time out of the reading program for children to read. When reading is regarded as an interruption, or an innovation, it suggests that the program of educating children to read has been widely off the mark. Although the materials and the processes they engage must be high caliber, the amount of engagement must be high for a reading program to be more than a name.

Although we have developed considerable sophistication about comprehension processes and children's interaction with materials, we are not yet close to a concept of a reading program that is sufficient to recognize one. We have not ruled out the child in the haystack reading a book (perhaps with adjunct questions) or the princess in the courtyard being privately tutored. While these may be the occasion for the acquisition of reading, they are so unusual as to fall outside the scope of activities known as a reading program. Reading programs are found in schools which are large, complex institutions with many sorts of individuals fulfilling a diversity of roles. It stands to reason that the characteristics of the institution, the individuals who occupy it, and the relationships among groups within it should condition the reading program.

Despite the unavoidable, social context in which learning to read occurs, the social structure in which reading programs are embedded has been excluded from educators' concepts of the programs. For example, the Right to Read system contracted the American Institute for Research to identify exemplary reading programs. A typical example of the 12 that were found to meet preestablished criteria was a Title I reading center program in Broward County, Florida, described as follows:

> Students from low income families who are functioning approximately 2 years below grade level in reading in grades 1-6 come to the main Reading Center or one of the 11 school-based centers for 1 hour of instruction every other day (a total of 5 hours every 2 weeks). Based on the results of staff-developed diagnostic tests, teachers at the centers prescribe an individualized program of learning and reinforcement activities for each child. A child spends part of each hour receiving direct instruction from the teacher in skills development and the remainder of the hour in reinforcement activities under the supervision of an aide. Staff of the Research Department of Broward County Schools work in close cooperation with program managers and instructors to determine the extent to which program objectives are realized, and to identify changes that should be made in the program (Bowers, Campeau, & Roberts, 1974, p. 62).

The United States Office of Education has attempted to compile and disseminate "educational programs that work." The Far West Laboratory for Educational Research and Development issued in the fall of 1978, a publication containing descriptions of these programs. One program, entitled "Learning to Read by Reading," took place in California and was claimed to be "a unique method of teaching reading to sub-par achievers at upper elementary through junior college levels who have failed to progress through the use of conventional methods and materials" (pp. 7-42). A description is given, including the kinds of symbols used, the length of stories, the number of books read orally, and the number of plays enacted during the course of the program. No statement was made about the social context, its structure, or the personal characteristics of the participants in the program.

We can be grateful that sociologists and anthropologists, who are not so immersed in reading programs as to make assumptions about their context, have reminded educators of the degree to which classrooms and schools may be regarded as minisocieties. For example, in a study led by anthropologist Jules Henry, innercity schools were observed by a team of investigators (Gouldner, 1978). They observed that

> Schooling was an experience that seemed to reward those children who most readily conformed to the social order established by the school bureaucracy. The schools were more bearable places for those children who allowed their individuality and autonomy to be shaped to the constraints of the classroom and the rest of the institution.
>
> Some of the children seemed to respond more willingly than others to the admonitions printed across the tops of the posters decorating the front wall of one first grade classroom: "Sit Nice and Straight," "Don't Talk," "Raise Your Hand," "Don't Chew Gum," "Do Your Best Work," "Take Care of Yourself," "Be Happy," "Smile." Although these slogans were not necessarily followed all the time by the high achievers who were the teacher's pets, they incorporated quite well how the teacher wished the students to behave most of the time. The stars participated in classroom activities, followed directions well, and gave correct answers, and although they were given much more freedom to bend the rules than the rest of the children, their teacher would curb them when they went too far. The pets were adept at responding to their teacher's ways of conducting her class and her need to keep order. On their own level, they possessed an almost uncanny grasp of how the bureaucracy worked (p. 85).

Cazden (this volume) suggests a number of social interaction variables that seem to be influential. For instance, the teacher's method of interacting is different with higher achieving and lower achieving reading groups within the classroom. The higher group receives more freedom, responsibility, and opportunity for self-initiated learning. The lower group is required to attend to questions and reading selections uniformly at the teacher's request. Children in the lower reading group are interrupted, guided, and corrected more frequently than those of the upper group (Allington, 1980). An analysis of these interruptions shows that they may inhibit, rather than enhance, learning. Doesn't this imply that because of the social structure created by the teacher in a classroom, the children's opportunities for cognitive interaction with written materials are substantially altered?

One dimension of the social structure is the group of children who are the learners. Their interactions with one another have been reported to be associated with learning by Walberg and his collaborators (this volume). They found that learning was positively associated with cohesiveness (how well students know each other). But it was negatively associated with friction (whether students are responsible for petty quarrels). Other negatively associated variables relating to interactions include apathy (students don't care what the class does) and competitiveness. They found paradoxically that difficulty of the work was positively associated with learning in high school but negatively associated with elementary school reading achievement. The power of some of these variables is surprising. For instance, cohesiveness and friction correlated to reading achievement at the level of .80 and .90. It seems vitally important to incorporate social goals into the reading program explicitly. To neglect them may be not only to subvert the academic intentions of the program, but also to condone a social system that produces dubious effects and consequences.

Another girder of the social structure of a reading program is the administration. In the current volume, Samuels points out that elements that have been common to exemplary reading programs emerging from six nationwide surveys include: leadership, teacher training, specific reading objectives, high intensity treatment, and efficient use of time. In addition, he noted that lack of leadership and inefficient use of time were common to unsuccessful efforts. These two variables seem to distinguish the remarkably effective programs from the apparently ineffective programs. The variable of leadership refers to enthusiasm, goal-setting, and organization of social and material relationships provided by a person in the reading program. It is noteworthy that this social variable has been identified solely through studies that are extremely applied with the premium on locating and disseminating effective systems. Bodies of research that are theoretically based, such as the one summarized by Jenkins and Pany, and are related to instruction have not accented this feature of the reading program.

The school is a unit in the social system of a community. Yet, the relationships between the school and the community have not been fully examined by reading education researchers, save for the recognition that "socioeconomic status" of pupils is an overpowering determinant of school achievement. In the present volume, two authors point to a linkage between the child in the reading program and the community, which is through the oral language. Cazden reports that in a highly effective Hawaiian reading program the cultural congruence of context for reading is high. The reason for this is that the classroom "lesson-talk" is similar to the collaborative narrative of personal experience in Hawaii. In other words, forms of spoken interaction, which entail social systems that occur in the community, also prevail in the classroom. As another instance, Elley illustrated that a highly effective reading and language program in bilingual communities in the South Pacific was developed by reading stories to children in the second language that were highly familiar to them in their first language. Although relationships between the school and its surrounding culture have not been articulated, they are implicit in our conception of a reading program. Tending and raising sheep is not part of the curriculum in most innercity American schools, although it is universal in rural schools in New Zealand which is a sheep raising country. Beyond the obvious, however, we have not expressed the values, knowledge, activities, and competencies important for the community that should be learned in schools nor examined the congruence between our aspirations and our practices in this regard.

A model of a reading program that may be sufficient might be described in terms of several categories of variables: student learning and interaction; reading materials; and a social system of teaching, administering, and community living. This book has focused primarily on the processes of learning and comprehension that must occur to say that a reading program exists. Cognitive interactions between the learner and the materials have been described with an attempt to identify variables in the materials that facilitate comprehension. Interactions between students who exist within the circle of learning seem to be important for the cognitive accomplishments that are the most widely shared goals of the reading instructional system. The social context for learning—which consists of school and community personnel, their behaviors, and their relationships to the learner—is crucial. Most teachers can recognize a program with these factors. Although researchers have succeeded in externalizing some important processes of comprehending, the available analysis of content is superficial, and our knowledge of the social context is primitive.

References

ALLINGTON, R. Teacher interruption behaviors during primary grade oral reading. *Journal of Educational Psychology*, 1980, *72* (3), 371-377.

BARTON, A., & WILDER, D. Research and practice in the teaching of reading: A progress report. In M.B. Miles (Ed.), *Innovation in education.* New York: Columbia University Bureau of Publications, 1964.

BECK, I., McKEOUEN, M., McCASLIN, E., & BURKES, A. *Instructional dimensions that may affect reading comprehension: Examples from two commercial reading programs.* Pittsburgh: Learning Research & Development Center, 1979.

BOND, G.L., & DYKSTRA, R. The cooperative research program in first grade reading instruction. *Reading Research Quarterly*, 1967, *2* (4), 5-142.

BOWERS, J.E., CAMPEAU, P.L., & ROBERTS, A.O.H. *Identifying, validating, and multi-media packaging of effective reading programs.* Palo Alto, Calif.: American Institutes for Research, 1974.

CHALL, J. *Learning to read: The great debate.* New York: McGraw-Hill, 1967.

FAR WEST LABORATORY FOR EDUCATIONAL RESEARCH AND DEVELOPMENT. *Education programs that work*. San Francisco: Far West Lab., 1978.

GOULDNER, H. *Teachers' pets, troublemakers, and nobodies: Black children in elementary school.* Westport, Conn.: Greenwood Press, 1978.

HARRIS, A., & SERWER, B. The CRAFT project: Instructional time in reading research. *Reading Research Quarterly,* 1966, *2*, 27-57.

ROSENSHINE, B., & BERLINER, D. Academic engaged time. *British Journal of Teacher Education,* 1978, *4*, 3-16.

RUDDELL, R., & CREWS, R. *Person to person, teacher's edition.* Boston: Allyn & Bacon, 1978.

STAUFFER, R. *Directing the reading-thinking process.* New York: Harper & Row, 1969.